



UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ENGENHARIA RURAL

**PROBABILIDADE E ESTATÍSTICA
APLICADA À HIDROLOGIA**

Rita Cabral Guimarães

2011

ESTATÍSTICA E PROBABILIDADE APLICADA À HIDROLOGIA

1. Introdução

Nenhum processo hidrológico é puramente determinístico, isto é, não é possível determinar com exactidão a realização desse processo, pois ele está sujeito à acção de factores aleatórios. Por exemplo, apesar de ser possível prever com alguma antecedência a ocorrência de precipitação, não é possível determinar qual a quantidade exacta de precipitação que irá ocorrer.

Este facto, parece estabelecer uma dificuldade básica no planeamento e gestão de qualquer sistema hidrológico, uma vez que para planear e gerir é fundamental conhecer o comportamento futuro dos processos que integram esse sistema hidrológico. No entanto, esta dificuldade pode ser ultrapassada, considerando que os processos hidrológicos são processos estocásticos, isto é, processos governados pelo menos em parte por factores aleatórios. Se são processos estocásticos podem ser tratados recorrendo às leis de probabilidade e à estatística, sendo possível determinar qual a probabilidade duma realização desses processos se situar dentro de determinados intervalos. Por exemplo, se chover, pode-se determinar com antecedência qual probabilidade de ocorrer um determinado valor de precipitação.

2. Distribuições de Frequência

Em estatística **população** designa um conjunto de elementos com alguma característica comum, por exemplo: os rios portugueses ou as precipitações anuais numa bacia hidrográfica. Pode-se dizer que a estatística se ocupa do estudo das propriedades das populações, populações estas que podem ser finitas ou infinitas conforme for finito ou infinito o número dos seus elementos. No entanto, e porque a observação de toda a população nem sempre é possível, o estudo das propriedades dessa população tem de ser feito sobre um seu subconjunto finito que se supõe ser representativo e se designa por **amostra**.

Quando, a partir da informação contida numa amostra, se tiram conclusões, expressas em termos de probabilidade, sobre toda a população entra-se no domínio da inferência estatística.

Considere-se uma amostra constituída por um determinado conjunto de dados x_1, x_2, \dots, x_n .

A diferença entre o maior e o menor dos valores dos dados, chama-se **amplitude dos dados**, I .

$$I = \text{maior } x_i - \text{menor } x_i \quad (2.1)$$

Para resumir grandes quantidades de dados é usual distribuí-los em **classes**. O número de indivíduos pertencentes a cada classe denomina-se **frequência absoluta** da classe. A razão entre a frequência absoluta da classe e a frequência total (número total de valores da amostra) chama-se **frequência relativa** da classe.

À distribuição dos dados em classes com as respectivas frequências absolutas, chama-se **distribuição de frequências** ou **distribuição empírica** e à distribuição dos dados em classes com as respectivas frequências relativas, chama-se **distribuição de frequências relativas** ou **distribuição das percentagens**. (ver Quadro 2.2. do exemplo 2.1.)

Geralmente, o **número de classes**, m , deverá ser entre 5 e 20, no entanto, pode utilizar-se, para cálculo do número de classes, a fórmula sugerida por STURGES:

$$m = 1 + \log_2 n = 1 + 3,32193 \log_{10} n \quad (2.2)$$

Determinado o número de classes e uma vez conhecida a amplitude dos dados I , a **amplitude de cada classe**, c , pode ser determinada por:

$$c = \frac{I}{m} \quad (2.3)$$

Exemplo 2.1

Considerem-se as precipitações anuais registadas na estação de Castro D'Aire durante 79 anos, apresentadas no Quadro 2.1.

A amplitude dos dados, determina-se facilmente pela equação (A2.1):

$$I = 3249,6 - 870,9 = 2378,7 \text{ mm},$$

o número de classes, utilizando a equação (2.2), é:

$$m = 1 + 3,32193 \log_{10} 79 = 7 \text{ classes,}$$

e a amplitude de cada classe, determina-se recorrendo à equação (2.3).

$$c = \frac{I}{m} = \frac{2378,7}{7} = 340 \text{ mm}$$

Isto é, a 1ª classe terá como limite inferior o valor 870,9 mm e como limite superior 1210,9 mm (870,9 + 340), a 2ª classe terá como limite inferior 1210,9 mm e como limite superior 1550,9 mm (1210,9 + 340), e assim sucessivamente até ao limite superior da última classe.

O apuramento dos valores pertencentes a cada classe que conduz às frequências absolutas e relativas de cada classe, não oferece qualquer dificuldade, considerando que um **determinado valor x pertence a uma classe** quando e só quando **é maior que o limite inferior e menor ou igual que o limite superior** dessa classe.

A divisão da amostra em classes bem como as frequências absolutas e relativas de cada classe são apresentadas no Quadro A2.2.

Quadro 2.1. Precipitação anual (mm) em Castro D'Aire

Ano	Precipitação (mm)	Precipitação ordenada de forma crescente (mm)	Nº de ordem
1916/17	2118,2	870,9	1
1917/18	1001,2	903,5	2
1918/19	2093,2	912,8	3
1919/20	1556,4	915,8	4
1920/21	1290,6	1001,2	5
1921/22	1785,4	1039,2	6
1922/23	1830,2	1055,4	7
1923/24	2150,2	1076,1	8
1924/25	1749,6	1127,1	9
1925/26	2221,6	1144,5	10
1926/27	2024,1	1180,0	11
1927/28	1923,7	1201,0	12
1928/29	1127,1	1239,6	13
1929/30	2630,9	1247,2	14
1930/31	1481,2	1254,0	15
1931/32	1461,0	1275,7	16
1932/33	1334,4	1290,6	17
1933/34	1301,2	1298,7	18
1934/35	1581,0	1300,3	19
1935/36	3249,6	1301,2	20
1936/37	2069,0	1334,4	21
1937/38	1254,0	1344,7	22
1938/39	1974,0	1392,9	23
1939/40	2059,6	1411,7	24
1940/41	2569,6	1422,9	25
1941/42	1520,6	1426,8	26
1942/43	1664,2	1432,0	27

1943/44	1344,7	1442,0	28
1944/45	915,8	1451,9	29
1945/46	1763,0	1461,0	30

Quadro 2.1. (Cont.) Precipitação anual (mm) em Castro D'Aire

Ano	Precipitação (mm)	Precipitação ordenada de forma crescente (mm)	Nº de ordem
1946/47	2079,3	1478,1	31
1947/48	1411,7	1481,2	32
1948/49	912,8	1496,4	33
1949/50	1201,0	1504,2	34
1950/51	1903,9	1520,6	35
1951/52	1625,0	1556,4	36
1952/53	1076,1	1567,9	37
1953/54	1275,7	1578,2	38
1954/55	1699,5	1581,0	39
1955/56	2150,9	1585,4	40
1956/57	1039,2	1588,2	41
1957/58	1588,2	1595,9	42
1958/59	1746,1	1603,3	43
1959/60	2563,6	1625,0	44
1960/61	1987,4	1664,2	45
1961/62	1585,4	1689,7	46
1962/63	1832,1	1699,5	47
1963/64	2201,2	1746,1	48
1964/65	1180,0	1749,6	49
1965/66	2806,9	1763,0	50
1966/67	1595,9	1785,4	51
1967/68	1422,9	1826,1	52
1968/69	2280,0	1830,2	53
1969/70	1496,4	1832,1	54
1970/71	1567,9	1903,9	55
1971/72	1300,3	1923,7	56
1972/73	1478,1	1930,1	57
1973/74	1689,7	1974,0	58
1974/75	1239,6	1987,4	59
1975/76	903,5	2000,2	60
1976/77	2314,0	2024,1	61
1977/78	2124,2	2059,6	62
1978/79	2599,1	2069,0	63
1979/80	1451,9	2079,3	64
1980/81	1144,5	2093,2	65
1981/82	1504,2	2118,2	66
1982/83	1426,8	2124,2	67
1983/84	1603,3	2150,2	68
1984/85	2000,2	2150,9	69
1985/86	1578,2	2201,2	70
1986/87	1392,9	2221,6	71
1987/88	1930,1	2280,0	72
1988/89	870,9	2314,0	73
1989/90	1432,0	2563,6	74
1990/91	1442,0	2569,6	75
1991/92	1055,4	2599,1	76
1992/93	1247,2	2630,9	77

1993/94	1826,1	2806,9	78
1994/95	1298,7	3249,6	79

Quadro 2.2. Distribuição de frequências e distribuição de frequências relativas da precipitação anual

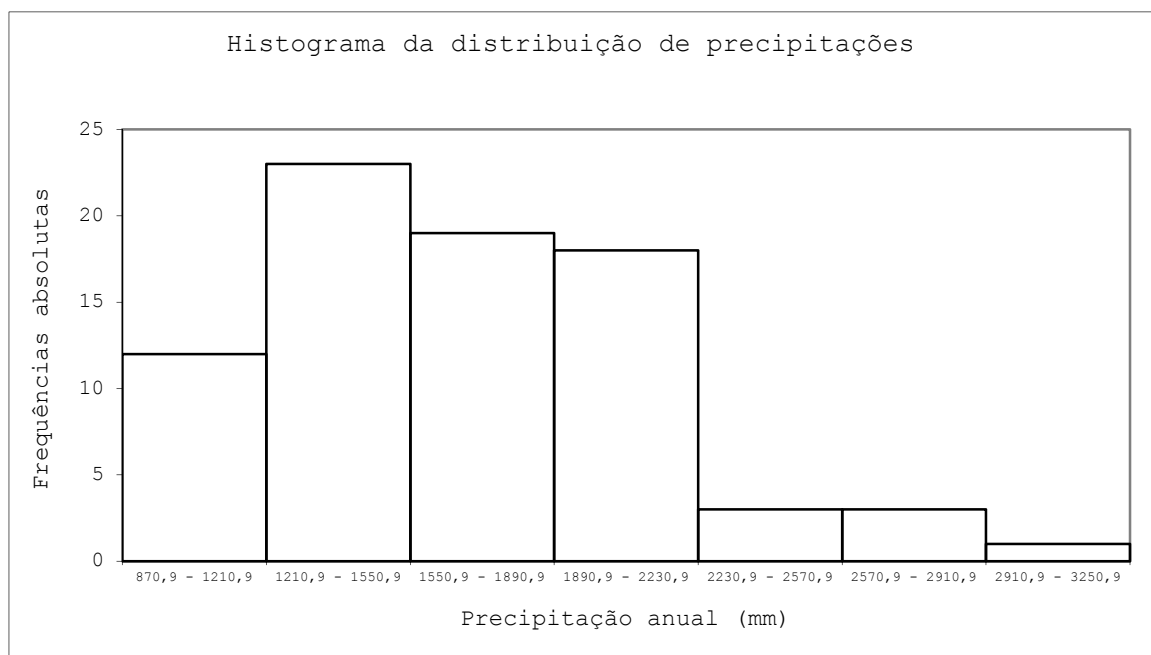
Classes de Precipitação Anual (mm)	Frequências absolutas	Frequências relativas
870,9 - 1210,9	12	$12/79 = 0,151898734$
1210,9 - 1550,9	23	$23/79 = 0,291139241$
1550,9 - 1890,9	19	$19/79 = 0,240506329$
1890,9 - 2230,9	18	$18/79 = 0,227848101$
2230,9 - 2570,9	3	$3/79 = 0,037974684$
2570,9 - 2910,9	3	$3/79 = 0,037974684$
2910,9 - 3250,9	1	$1/79 = 0,012658228$
TOTAL	79	1

A **representação gráfica** numa distribuição de frequências, fornece uma visão global da distribuição. Esta representação gráfica pode ser feita através de um **histograma**.

O **histograma** é uma sucessão de rectângulos adjacentes, tendo cada um deles por base um segmento que corresponde à amplitude de cada classe e por altura as respectivas frequências absolutas ou relativas.

Na figura 2.1. apresenta-se o histograma das frequências absolutas referente ao exemplo 2.1.

Figura 2.1. Histograma das frequências absolutas para a precipitação anual em Castro D'Aire.



3. Localização, Dispersão e assimetria

Aspectos fundamentais para a caracterização das distribuições de frequência são a localização, dispersão e assimetria.

Far-se-á distinção entre as grandezas avaliadas a partir da população - **parâmetros** - das grandezas calculadas com base na amostra - **estatísticas** -. Assim, os parâmetros serão representados por letras gregas ($\mu, \sigma, \gamma, \dots$) e as estatísticas serão representadas por letras latinas (\bar{x}, S, g, \dots).

3.1 Localização

a) Média

O mais importante parâmetro de localização é a **média** ou valor médio e representa o centro de gravidade do sistema.

A média de uma amostra constituída por x_1, x_2, \dots, x_n , designa-se por \bar{x} e para dados não classificados, define-se por,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

Para dados classificados, utiliza-se, para cálculo da média, expressão,

$$\bar{x} = \frac{\sum_{j=1}^m \eta_j x'_j}{n} = \sum_{j=1}^m f_j x'_j \quad (3.2)$$

onde m é número de classes, η_j representa a frequência absoluta de cada classe, x'_j o ponto médio de cada classe e f_j a frequência relativa de cada classe. A esta última expressão dá-se o nome de **média ponderada**.

b) Mediana

Considere-se uma amostra, x_1, x_2, \dots, x_n , a **mediana**, pode definir-se como o valor central da amostra ordenada por ordem crescente, $x_1 \leq x_2 \leq \dots \leq x_n$.

Assim, a mediana, M , pode definir-se por duas expressões:

- Se a amostra tem número ímpar de dados, $n = 2k + 1$, e a mediana vem,

$$M = x_{k+1} \quad \text{para } n = 2k + 1 \quad (3.3)$$

- Se a amostra tem número par de dados, $n = 2k$, e a mediana vem,

$$M = \frac{x_k + x_{k+1}}{2} \quad \text{para } n = 2k \quad (3.4)$$

c) Moda

A **moda** de uma amostra x_1, x_2, \dots, x_n , define-se como o valor mais frequente da amostra. É a medida de localização menos usada em hidrologia, pois em amostras de dados hidrológicos (precipitações, caudais, etc) é pouco provável que haja valores exactamente iguais. No entanto para cálculo da moda, Mod , pode utilizar-se a expressão,

$$Mod = \bar{x} - 3(\bar{x} - M) \quad (3.5)$$

Exemplo 3.1 Cálculo da média, mediana e moda

Considerem-se as precipitações anuais registadas na estação de Castro D'Aire (Quadro 2.1.) e a respectiva distribuição de frequências e distribuição de frequências relativas (Quadro 2.2.)

a) cálculo da média

Utilizando a equação (3.1) vem,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2118,2 + \dots + 1298,7}{79} = 1672,5mm$$

A média ponderada determina-se utilizando a equação (A3.2),

$$\bar{x} = \frac{\sum_{j=1}^m \eta_j x'_j}{n} = \frac{\left(\frac{870,9 + 1210,9}{2} \times 12 \right) + \dots + \left(\frac{2910,9 + 3250,9}{2} \times 1 \right)}{79} = 1677,9mm$$

$$\bar{x} = \sum_{j=1}^m f_j x'_j = \left(\frac{870,9 + 1210,9}{2} \times 0,1519 \right) + \dots + \left(\frac{2910,9 + 3250,9}{2} \times 0,1266 \right) = 1677,9 \text{ mm}$$

b) cálculo da mediana

Uma vez que n é ímpar, utiliza-se a equação (A3.3) para cálculo da mediana,

$$n = 79 \Rightarrow k = \frac{n-1}{2} = \frac{78}{2} = 39$$

$$M = x_{k+1} = x_{40} = 1585,4 \text{ mm}$$

b) cálculo da moda

Utilizando a expressão (3.5), vem,

$$Mod = \bar{x} - 3(\bar{x} - M) = 1672,5 - 3(1672,5 - 1585,4) = 1411,2 \text{ mm}.$$

Para esta distribuição de precipitações, tem-se que,

$$\boxed{\bar{x} > M > Mod}$$

3.2 Dispersão

A dispersão pode definir-se como a posição dos dados em relação a uma referência fixa. Quando esta referência é a média, a dispersão indica o modo como os dados se espalham à volta do valor médio.

a) Desvio Padrão

Uma importante medida de dispersão é o **desvio padrão** que mostra o comportamento do conjunto de desvios em relação à média. Se a dispersão é grande, os desvios dos dados em relação à média são grandes e o desvio padrão será elevado. O contrário também se verifica quando os desvios são pequenos.

O desvio padrão de uma amostra constituída por x_1, x_2, \dots, x_n , designa-se por S e para dados não classificados, define-se por,

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.6)$$

Para dados classificados, vem,

$$S = +\sqrt{\frac{\sum_{j=1}^m \eta_j (x'_j - \bar{x})^2}{n}} \quad (3.7)$$

Quando as amostras são pequenas, utiliza-se o **desvio padrão corrigido**,

$$S = +\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3.8)$$

e

$$S = +\sqrt{\frac{\sum_{j=1}^m \eta_j (x'_j - \bar{x})^2}{n-1}} \quad (3.9)$$

Ao quadrado do desvio padrão, chama-se **variância**, s^2 , e para amostras pequenas vem,

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (3.10)$$

b) Desvio Médio

Outra forma de analisar o conjunto de desvios em relação à média é considerar o módulo dos desvios. Isto conduz ao conceito de **desvio médio**, d , onde os desvios perdem o sinal, e quanto maior o valor do desvio médio, mais as observações se afastam da média da amostra. O desvio médio de uma amostra x_1, x_2, \dots, x_n determina-se, para dados não classificados, por,

$$d = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (3.11)$$

e para dados classificados por,

$$d = \frac{\sum_{j=1}^m \eta_j |x'_j - \bar{x}|}{n} \quad (3.12)$$

b) Coeficiente de Variação

É um parâmetro adimensional que mede a variabilidade da amostra e define-se por,

$$C_v = \frac{S}{\bar{x}} \times 100\% \quad (3.13)$$

Quanto maior o coeficiente de variação, maior é o desvio padrão em relação à média, isto é, mais dispersos estão os dados em torno da média.

d) Variável Reduzida

A **variável reduzida**, z , mede o desvio, de cada observação da amostra x_1, x_2, \dots, x_n , em relação à média em unidades de desvio padrão. É, portanto, uma quantidade abstracta independente das unidades usadas.

$$z_i = \frac{x_i - \bar{x}}{S} \quad (3.14)$$

Assim, o total de variáveis reduzidas da amostra, de $i = 1, 2, \dots, n$, apresenta média nula e desvio padrão igual à unidade. Isto é,

$$\bar{z} = \frac{\sum_{i=1}^n z_i}{n} = 0,0 \quad (3.15)$$

$$S_z = +\sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}} = 1,0 \quad (3.16)$$

Exemplo 3.2 Cálculo do desvio padrão, desvio médio, coeficiente de variação e variável reduzida

Considerem-se as precipitações anuais registadas na estação de Castro D'Aire (Quadro 2.1.) e a respectiva distribuição de frequências e distribuição de frequências relativas (Quadro 2.2.)

a) cálculo do desvio padrão

Utilizando a equação (A3.8) vem,

$$S = +\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = +\sqrt{\frac{(2118,2 - 1672,5)^2 + \dots + (1298,7 - 1672,5)^2}{79 - 1}} = 479,4mm$$

e utilizando a equação (A3.9) vem,

$$S = +\sqrt{\frac{\sum_{j=1}^m \eta_j (x'_j - \bar{x})^2}{n - 1}} = +\sqrt{\frac{12 \times (1040,9 - 1672,5)^2 + \dots + 1 \times (3080,9 - 1672,5)^2}{79 - 1}} = 460,0mm$$

b) cálculo do desvio médio

Utilizando a equação (A3.11) vem,

$$d = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{|2118,2 - 1672,5| + \dots + |1298,7 - 1672,5|}{79} = 381,8mm$$

e utilizando a equação (A3.12),

$$d = \frac{\sum_{j=1}^m \eta_j |x'_j - \bar{x}|}{n} = \frac{12 \times |1040,9 - 1672,5| + \dots + 1 \times |3080,9 - 1672,5|}{79} = 367,0mm$$

c) cálculo do coeficiente de variação

Utilizando a equação (3.13) vem,

$$C_v = \frac{S}{\bar{x}} \times 100 = \frac{479,4}{1672,5} \times 100 = 28,7\%$$

d) cálculo da variável reduzida

Utilizando a equação (3.14) as variáveis reduzidas de cada uma das observações da precipitação anual, são as apresentadas no Quadro 3.1. O valor médio e o desvio padrão foram calculados pelas equações (3.15) e (3.16), respectivamente.

Quadro 3.1. Variáveis reduzidas da precipitação anual em Castro D'Aire

Ano	Precipitação (mm)	Z_i
1916/17	2118,2	0,9
1917/18	1001,2	-1,4
1918/19	2093,2	0,9
1919/20	1556,4	-0,2
1920/21	1290,6	-0,8
1921/22	1785,4	0,2
1922/23	1830,2	0,3
1923/24	2150,2	1,0
1924/25	1749,6	0,2
1925/26	2221,6	1,1
1926/27	2024,1	0,7
1927/28	1923,7	0,5
1928/29	1127,1	-1,1
1929/30	2630,9	2,0
1930/31	1481,2	-0,4
1931/32	1461,0	-0,4
1932/33	1334,4	-0,7
1933/34	1301,2	-0,8
1934/35	1581,0	-0,2
1935/36	3249,6	3,3
1936/37	2069,0	0,8
1937/38	1254,0	-0,9
1938/39	1974,0	0,6
1939/40	2059,6	0,8
1940/41	2569,6	1,9
1941/42	1520,6	-0,3
1942/43	1664,2	0,0
1943/44	1344,7	-0,7
1944/45	915,8	-1,6
1945/46	1763,0	0,2
1946/47	2079,3	0,8
1947/48	1411,7	-0,5
1948/49	912,8	-1,6
1949/50	1201,0	-1,0
1950/51	1903,9	0,5
1951/52	1625,0	-0,1
1952/53	1076,1	-1,2
1953/54	1275,7	-0,8
1954/55	1699,5	0,1
1955/56	2150,9	1,0
1956/57	1039,2	-1,3
1957/58	1588,2	-0,2

1958/59	1746,1	0,2
1959/60	2563,6	1,9
1960/61	1987,4	0,7
1961/62	1585,4	-0,2
1962/63	1832,1	0,3
1963/64	2201,2	1,1
1964/65	1180,0	-1,0

Quadro 3.1. (Cont.) Variáveis reduzidas da precipitação anual em Castro D'Aire

Ano	Precipitação (mm)	Z_i
1965/66	2806,9	2,4
1966/67	1595,9	-0,2
1967/68	1422,9	-0,5
1968/69	2280,0	1,3
1969/70	1496,4	-0,4
1970/71	1567,9	-0,2
1971/72	1300,3	-0,8
1972/73	1478,1	-0,4
1973/74	1689,7	0,0
1974/75	1239,6	-0,9
1975/76	903,5	-1,6
1976/77	2314,0	1,3
1977/78	2124,2	0,9
1978/79	2599,1	1,9
1979/80	1451,9	-0,5
1980/81	1144,5	-1,1
1981/82	1504,2	-0,4
1982/83	1426,8	-0,5
1983/84	1603,3	-0,1
1984/85	2000,2	0,7
1985/86	1578,2	-0,2
1986/87	1392,9	-0,6
1987/88	1930,1	0,5
1988/89	870,9	-1,7
1989/90	1432,0	-0,5
1990/91	1442,0	-0,5
1991/92	1055,4	-1,3
1992/93	1247,2	-0,9
1993/94	1826,1	0,3
1994/95	1298,7	-0,8
Média	1672,5	0,0
Desvio Padrão	479,4	1,0

3.3 Assimetria

Assimetria é o grau de desvio, ou afastamento da simetria, de uma distribuição.

Quando se trabalha com distribuições de frequências, a assimetria pode ser estudada considerando a posição relativa dos três

parâmetros de localização: média, mediana e moda. Assim, nas distribuições **simétricas** (Figura 3.1), estes três parâmetros coincidem. Nas distribuições **assimétricas positivas** (desviadas para a direita) (Figura 3.2), **média>mediana>moda** e nas distribuições **assimétricas negativas** (desviadas para a esquerda) (Figura 3.3), **média<mediana<moda**.

A assimetria avalia-se pelo **coeficiente de assimetria**, g , sendo o valor deste coeficiente positivo nos desvios para a direita e negativo nos desvios para a esquerda.

$$g = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)S^3} \quad (3.17)$$

Figura 3.1. Distribuição simétrica

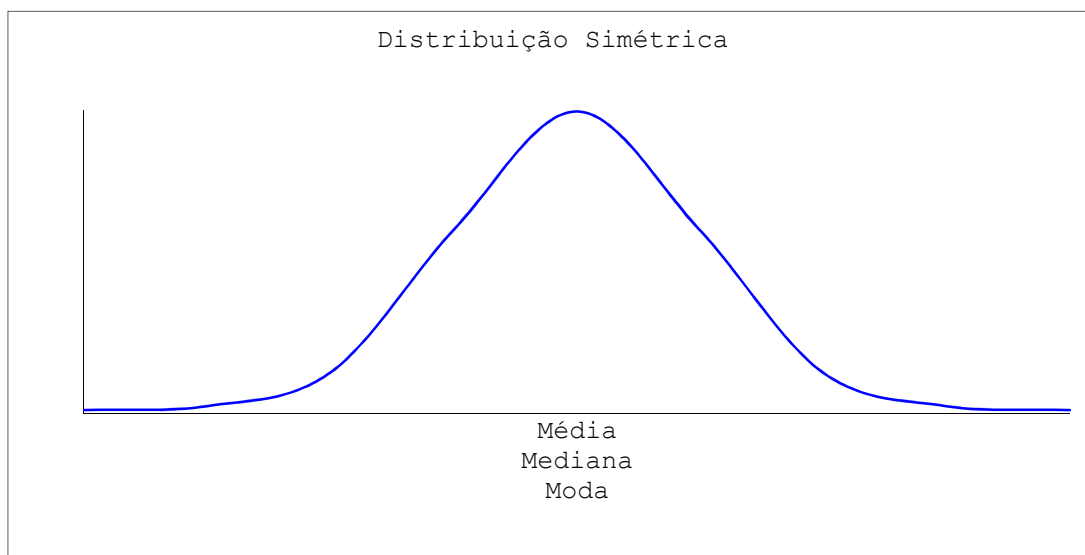


Figura 3.2. Distribuição assimétrica positiva

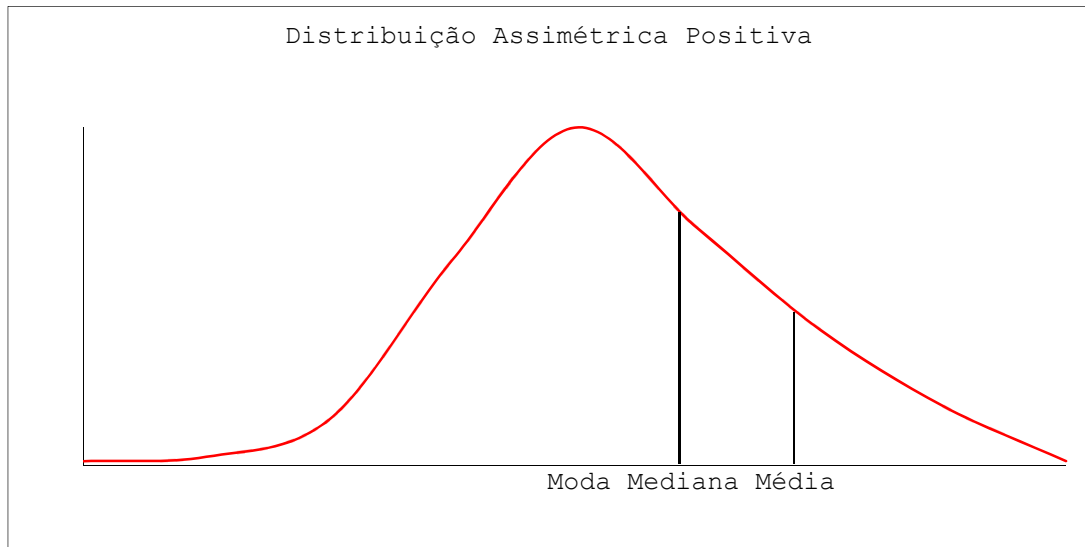
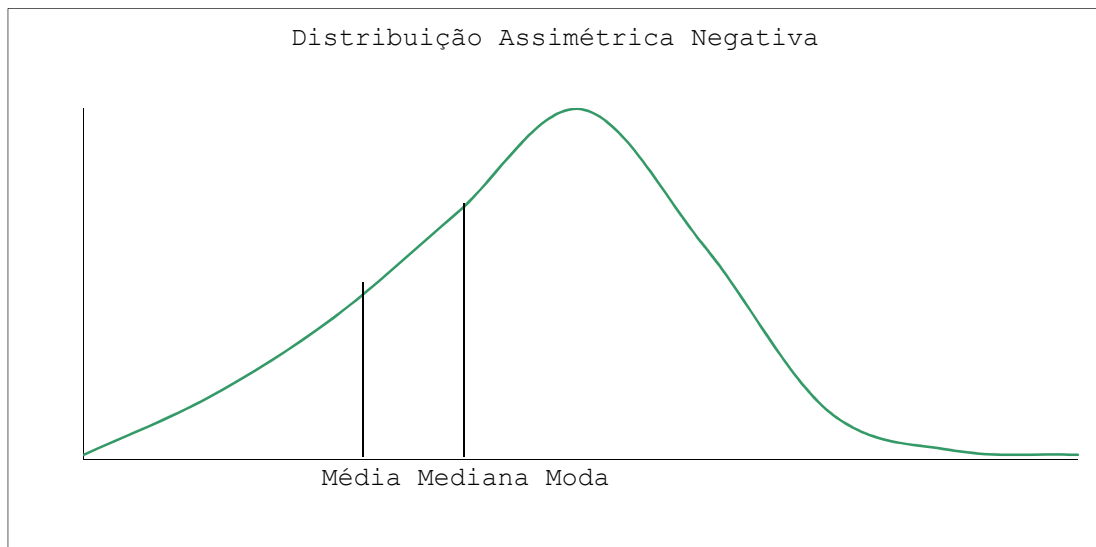


Figura 3.3. Distribuição assimétrica negativa

**Exemplo 3.3 Cálculo do coeficiente de assimetria**

Considerando as precipitações anuais registradas na estação de Castro D'Aire (Quadro A2.1.), o coeficiente de assimetria vem,

$$g = \frac{79[(2118,2 - 1672,5)^3 + \dots + (1298,7 - 1672,5)^3]}{(79 - 1)(79 - 2) \times 479,4^3} = 0,7$$

Como a distribuição tem assimetria positiva, significa que $\bar{x} > M > Mod$ (já determinado no exemplo 3.1), isto é, trata-se de uma distribuição desviada para a direita.

4. Distribuições de Probabilidade

4.1 Variável aleatória. Função de distribuição

Chama-se **variável aleatória** X a toda a variável susceptível de tomar diferentes valores de x^1 aos quais é possível afectar uma probabilidade. **Processo estocástico** é aqui entendido como uma colecção ordenada de variáveis aleatórias $X_1, X_2, X_3, \dots, X_n$ e onde a sucessão cronológica $x_1, x_2, x_3, \dots, x_n$ resultante da sua observação, representa uma única realização do processo.

Uma variável aleatória diz-se **discreta** se só pode tomar um número finito de valores, por exemplo: o número de dias com chuva numa semana, mês ou ano, ou o número de vezes que o caudal ultrapassou determinado valor. Uma variável aleatória diz-se **contínua** se pode assumir qualquer valor dentro de um determinado intervalo de números reais, por exemplo: a precipitação anual, a temperatura média diária, etc., podem tomar qualquer valor dentro de um certo intervalo limitado por um mínimo e por um máximo.

Sendo X uma variável aleatória, dá-se o nome de **função de distribuição** (ou função de distribuição de probabilidade) da variável X à função,

$$F(x) = P(X \leq x) \quad (4.1)$$

que representa a probabilidade de a variável aleatória X assumir um valor inferior ou igual a x .

Facilmente se verifica que a função de distribuição satisfaz as seguintes propriedades:

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x) \quad (4.2)$$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) \quad (4.3)$$

Quando X é uma **variável aleatória discreta**, só pode tomar valores dentro de um conjunto finito $\{x_1, x_2, \dots\}$. Neste caso o comportamento da variável aleatória é definido pela sua **função massa**

¹ Para evitar confusões, a variável aleatória representa-se por maiúsculas, X , e as observações (ou realizações) dessa variável por minúsculas, x .

de probabilidade (f.m.p.), que associa uma probabilidade a cada valor que a variável pode assumir:

$$\begin{aligned} P(x) &= 0 && \text{se } x \notin \{x_1, x_2, \dots\} \\ P(x_1) &= P(X = x_1) = P_1 \\ P(x_2) &= P(X = x_2) = P_2 \\ &\text{etc.} \end{aligned} \quad (4.4)$$

e pela sua **função de distribuição acumulada**, ou **função de distribuição** (f.d.):

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(x_i) \quad \text{com } i = 1, 2, \dots \quad (4.5)$$

Quanto à probabilidade de uma variável discreta X tomar um valor compreendido num intervalo (a, b) pode ser determinada aplicando a equação (4.3),

$$P(a < X \leq b) = F(b) - F(a) = \sum_{x_i \leq b, x_i > a} P(x_i) \quad (4.6)$$

Exemplo 4.1 Função massa de probabilidade e função de distribuição para uma variável discreta

Exemplificando, se para um determinado período de tempo o número de dias nublados (X), em dado local, tiver a seguinte função massa de probabilidade (f.m.p.):

$$\begin{aligned} &0,15 \text{ se } x = 0 \\ &0,20 \text{ se } x = 1 \\ P(x) &= 0,30 \text{ se } x = 2 \\ &0,20 \text{ se } x = 3 \\ &0,15 \text{ se } x = 4 \end{aligned}$$

a função de distribuição (f.d.) será:

$$F(x) = \begin{array}{ll} 0 & \text{se } x < 0 \\ 0,15 & \text{se } 0 \leq x < 1 \\ 0,35 & \text{se } 1 \leq x < 2 \\ 0,65 & \text{se } 2 \leq x < 3 \\ 0,85 & \text{se } 3 \leq x < 4 \\ 1,00 & \text{se } 4 \leq x \end{array}$$

Nas figuras 4.1 e 4.2 estão representados os gráficos da f.m.p e f.d. para o número de dias nublados.

Quando X é uma **variável aleatória contínua** o seu comportamento é definido pela sua **função densidade de probabilidade** (f.d.p.), que define a probabilidade média da variável aleatória se situar dentro de um determinado intervalo.

Figura 4.1 - Função massa de probabilidade para a variável X

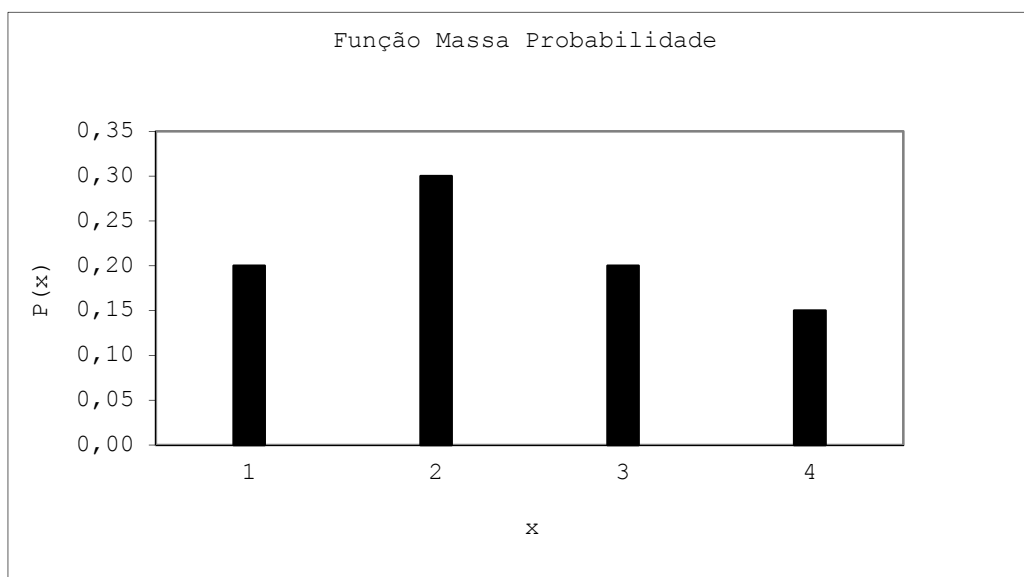
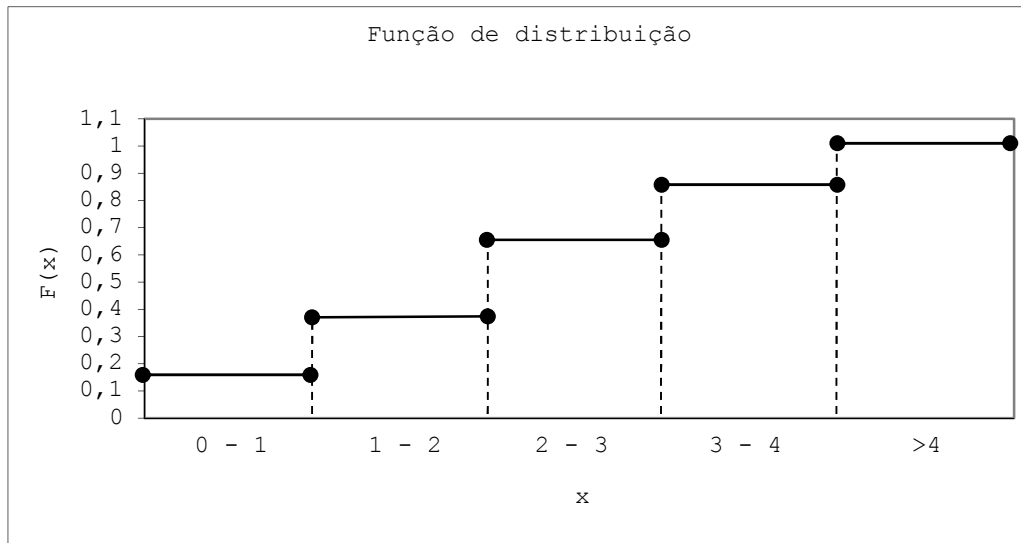


Figura 4.2 - Função de distribuição para a variável X



Considerando que no intervalo $(-\infty, x]$ a quantidade de probabilidade é $F(x)$, então no intervalo $(x, x + \Delta x]$, a quantidade de probabilidade será, de acordo com a equação (A4.3), $F(x + \Delta x) - F(x)$, e o quociente, $\frac{F(x + \Delta x) - F(x)}{\Delta x}$, será a quantidade de probabilidade média nesse intervalo. O limite,

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x) = \frac{dF(x)}{dx} \quad (4.7)$$

se existir, representa a densidade de probabilidade, $f(x)$. Onde resulta que a **função de distribuição** (f.d.) de uma variável contínua será,

$$F(x) = \int_{-\infty}^x f(x) dx \quad (4.8)$$

Note-se que, enquanto no caso discreto a f.m.p. é uma probabilidade $P(x) = P(X = x)$, no caso contínuo, a f.d.p. $f(x)$ é a densidade de probabilidade, nunca a probabilidade no ponto x .

Para variáveis contínuas:

$$f(x) > 0 \quad (4.9)$$

e

$$\int_{-\infty}^{+\infty} f(x) d(x) = 1 \quad (A4.10)$$

Quanto à probabilidade de a variável contínua X tomar um valor compreendido num intervalo (a, b) pode ser determinada por:

$$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx \quad (\text{A4.11})$$

e a probabilidade de X tomar um valor particular, a , é nula, visto que:

$$\int_a^a f(x)d(x) = 0 \quad (4.12)$$

Exemplo 4.2 Função densidade de probabilidade e função de distribuição para uma variável contínua.

A precipitação anual em Évora é uma variável aleatória contínua, X , com função de distribuição, $F(x)$, e função densidade de probabilidade, $f(x)$, dadas por,

$$F(x) = \int_{-\infty}^x f(x) dx$$

$$f(x) = \frac{1}{\sqrt{2\pi}203,5} \exp\left\{-\frac{(x - 651,8)^2}{2(203,5)^2}\right\}$$

Nas Figuras 4.3 e 4.4 mostra-se a representação gráfica destas duas funções.

Figura 4.3. Função densidade de probabilidade da variável X

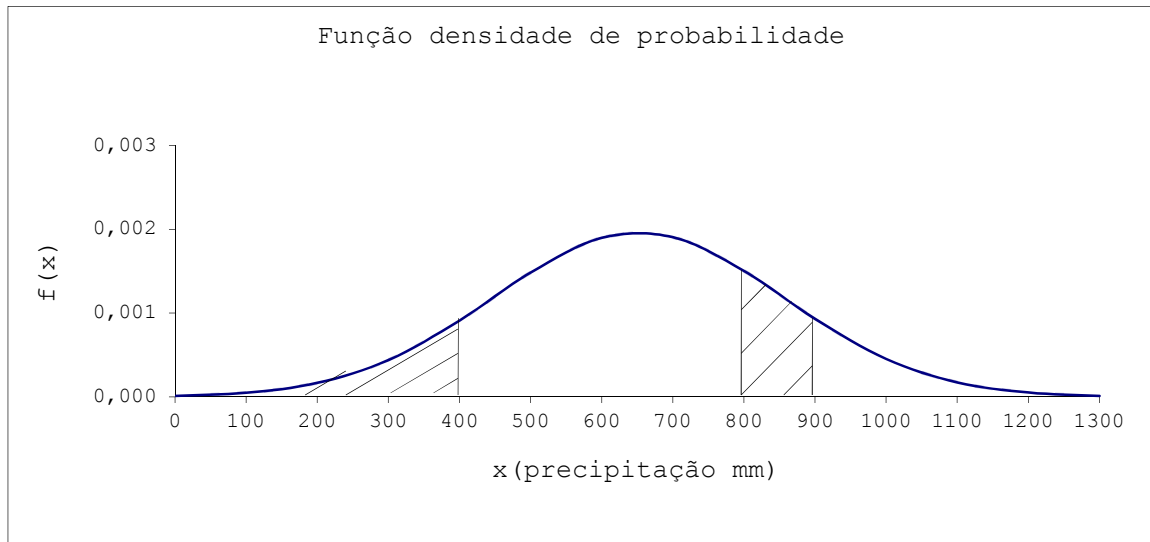
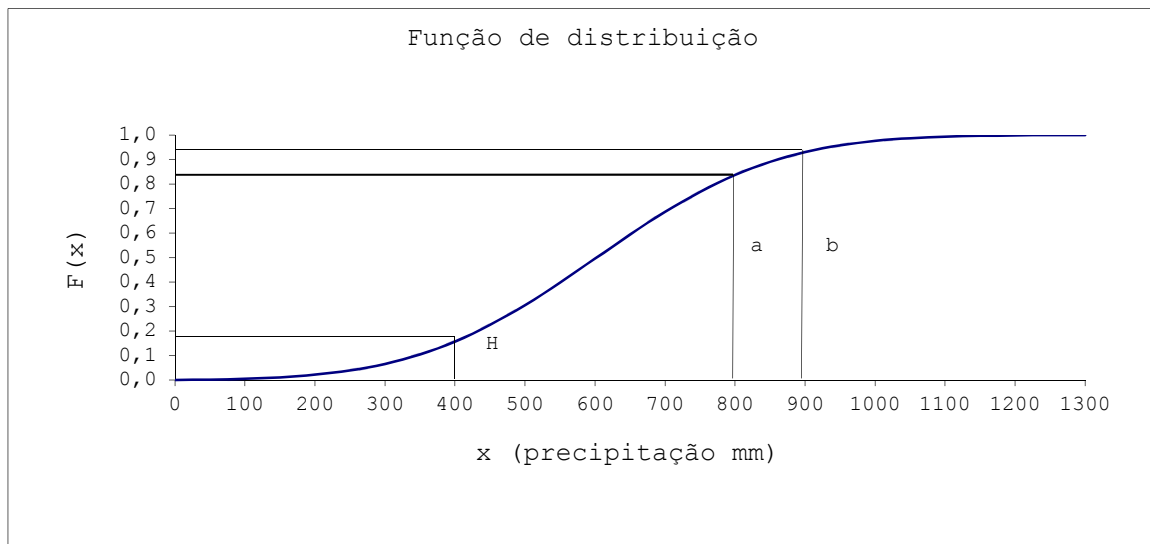


Figura 4.4- Função de distribuição da variável X



Na Figura 4.3, a área limitada pela curva e pelo eixo dos x é igual a um. A área a tracejado, correspondente às verticais $x = 0$ e $x = 400\text{mm}$, representa a probabilidade da precipitação em determinado ano ser igual ou menor que 400mm . A área a tracejado, correspondente às verticais $x = 800\text{mm}$ e $x = 900\text{mm}$, representa a probabilidade da precipitação tomar um valor entre 800 e 900 mm.

Na Figura 4.4. a altura H , corresponde à probabilidade da precipitação em determinado ano ser igual ou menor que 400mm . A altura $b - a$, representa a probabilidade da precipitação tomar um valor entre 800 e 900 mm.

5. Distribuições Teóricas

Existem muitas **distribuições teóricas**, que servem como **modelo probabilístico** de variáveis ou fenômenos aleatórios. Considerando que as variáveis hidrológicas são aleatórias, então elas podem ser representadas por algum tipo de distribuição teórica.

Apresentam-se de seguida as distribuições teóricas mais utilizadas em hidrologia.

5.1 Distribuições Discretas

a) Distribuição Binomial

A **distribuição Binomial** é o modelo probabilístico indicado para descrever o **número de sucessos** em repetidas provas de Bernoulli.

As provas de Bernoulli (ou experiências de Bernoulli) são sucessões de experiências aleatórias independentes, onde em cada uma delas só existem dois resultados possíveis: realização de determinado acontecimento e realização do contrário desse acontecimento. Considerando um qualquer acontecimento, A , de probabilidade $P(A) = p$, a realização de, A , diz-se "sucesso", a realização do contrário, \bar{A} , que tem probabilidade $P(\bar{A}) = 1 - p$, diz-se "insucesso".

Por exemplo, a ocorrência de precipitação em determinado dia do futuro, só tem dois resultados possíveis: ou chove (sucesso) ou não chove (insucesso) nesse dia. Então, a probabilidade de chover é p , e a probabilidade de não chover, será logicamente $1-p$.

Se a **variável aleatória**, X , designar o **número de sucessos** em n provas, diz-se que tem **distribuição Binomial** e escreve-se simbolicamente $B(n, p)$. A sua função massa de probabilidade é,

$$P(x) = P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n \quad (5.1)$$

e a sua função de distribuição é,

$$F(x) = \sum_{x_i \leq x} \frac{n!}{x_i!(n-x_i)!} p^{x_i} (1-p)^{n-x_i} \quad (5.2)$$

Exemplo 5.1

Considerando que em determinado rio ocorre uma cheia por ano e que a probabilidade desta cheia ser catastrófica é 10%, qual é a **probabilidade de ocorrência** de 3 destas cheias nos próximos 15 anos?

Neste caso, tem-se,

$$n = 15 \text{ anos}$$

$$x = 3$$

$$p = 0,1$$

logo, pela equação (5.1) vem,

$$P(3) = P(X = 3) = \frac{15!}{3! (15 - 3)!} 0,1^3 (1 - 0,1)^{15-3} = 0,1285$$

Isto é, nos próximos 15 anos a probabilidade de ocorrência de 3 cheias catastróficas neste rio é de 12,85%.

5.2 Distribuições Contínuas

a) Distribuição Normal

A mais importante e mais divulgada distribuição contínua de probabilidade é sem dúvida a distribuição Normal. Teoricamente, a função de distribuição da soma de n variáveis aleatórias tende para a distribuição Normal quando n aumenta indefinidamente, qualquer que seja a função de distribuição de cada uma das variáveis aleatórias. Por esta razão a distribuição Normal adapta-se bem a um grande número de variáveis hidrológicas, nomeadamente a precipitação anual e o escoamento anual, resultantes da soma de um grande número de variáveis aleatórias.

Uma variável aleatória X com uma função densidade de probabilidade,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < +\infty \quad (5.3)$$

diz-se que tem **distribuição Normal** com parâmetros μ e σ , e escreve-se simbolicamente, $N(\mu, \sigma)$. Os parâmetros μ e σ , são determinados por,

$$\mu = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{e} \quad \sigma = S = +\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$$

A sua função de distribuição é dada por,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (5.4)$$

Para se efectuar o estudo da distribuição Normal é necessário passar à **distribuição Normal Reduzida**, visto que os valores da função densidade de probabilidade e de distribuição são dados através de tabelas em função dos valores reduzidos. Isto consegue-se fazendo uma mudança de variável de modo a que a nova variável tenha valor médio igual a zero e desvio padrão igual à unidade. Isto é, transforma-se a variável X com $N(\mu, \sigma)$ numa variável Z com $N(0,1)$. Z é a variável reduzida, e é dada por,

$$Z = \frac{X - \mu}{\sigma} \quad (5.5)$$

Ao realizar-se esta transformação, **estandardiza-se a variável X** e neste caso a sua função densidade de probabilidade é,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < +\infty \quad (5.6)$$

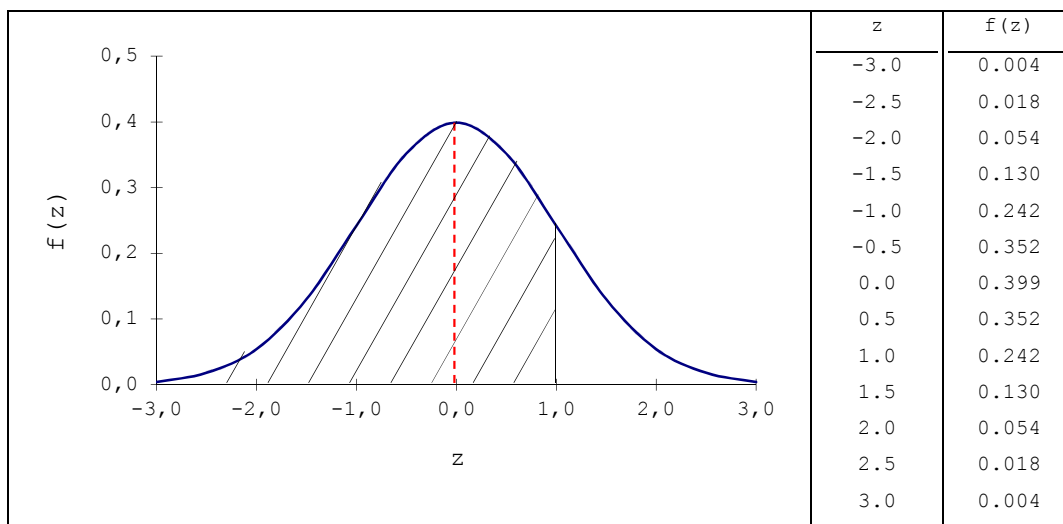
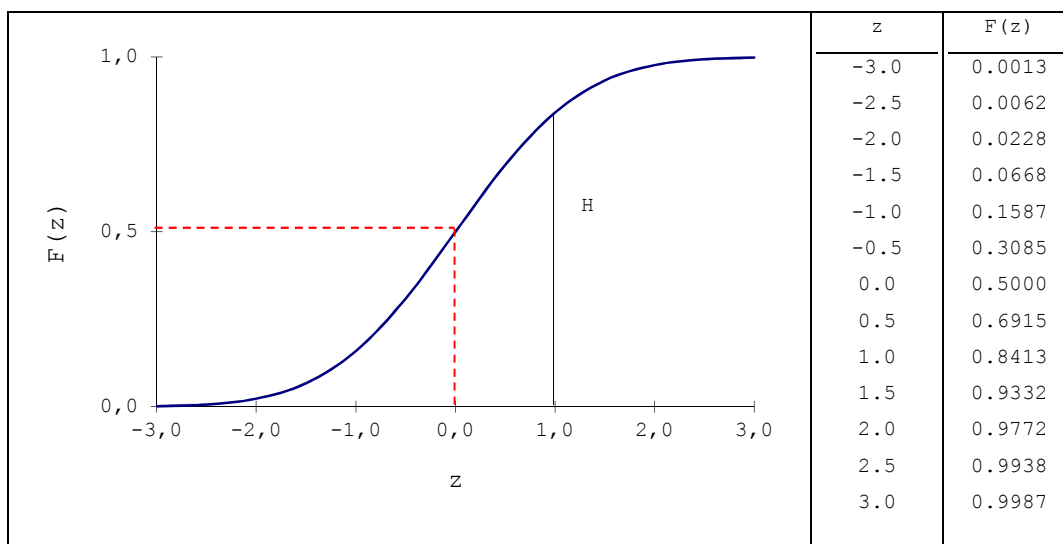
e a sua função de distribuição,

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \quad (5.7)$$

Os valores de $f(z)$ e $F(z)$ são dados por tabelas em função de z . Ver tabelas para a distribuição Normal, apresentadas no ponto 9.

Na figura 5.1 apresentam-se o gráfico da função densidade, $f(z)$, bem como os valores das ordenadas para os respectivos valores de z e na figura 5.2 o gráfico da função distribuição, $F(z)$. A altura H na figura 5.2 é a probabilidade acumulada correspondente à área tracejada na figura 5.1.

Fig. 5.1. Função densidade probabilidade $N(0,1)$

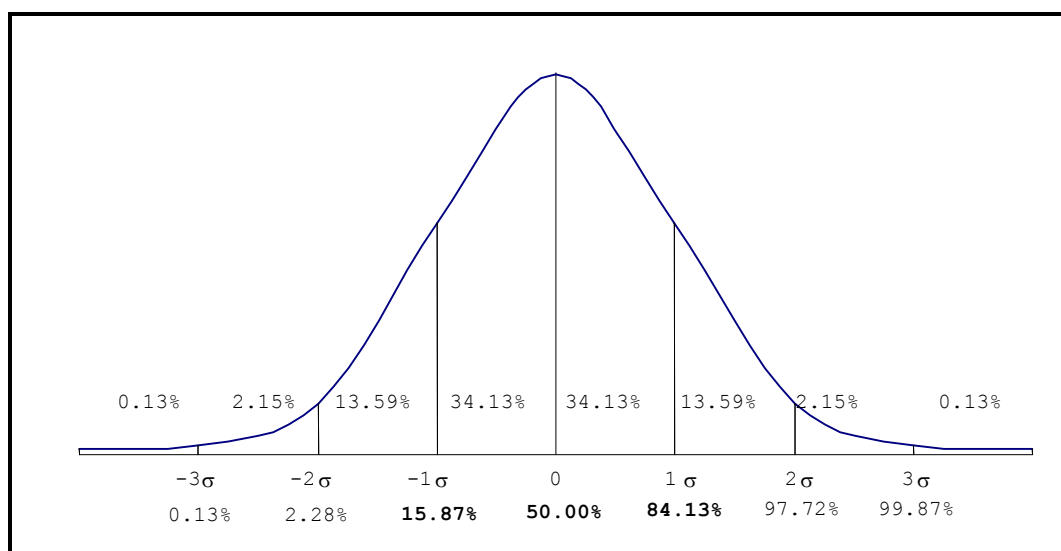
Fig. 5.2. Função de distribuição probabilidade $N(0,1)$ 

Como se pode verificar pelas Figuras 5.1 e 5.2, a distribuição Normal é uma **distribuição simétrica**, isto é caracteriza-se por ter uma densidade de probabilidade simétrica em relação à média, que é ao mesmo tempo mediana e moda. Isto significa que a probabilidade média que a variável aleatória tem de se situar no intervalo $(-\infty, \mu]$ é igual à probabilidade média que ela tem de se situar no intervalo $[\mu, +\infty)$.

Como se pode observar na figura 5.3 e comprovar com as tabelas (Ver 9), a área total limitada pela curva e pelo eixo dos x é unitária (100%). Também se pode observar que 50% da distribuição Normal 50% correspondem ao intervalo $(-\infty, \mu]$. Isto significa que a probabilidade

média de a variável aleatória tem de se situar no intervalo $(-\infty, \mu]$ é igual 50%. Ou, por outras palavras, a probabilidade média de a variável aleatória ser igual ou inferior ao valor médio é 50%. Também se pode observar na figura que 68.26% da distribuição normal correspondem ao intervalo $[\mu \pm \sigma]$, que 95.44% correspondem ao intervalo $[\mu \pm 2\sigma]$ e que 99.74% correspondem ao intervalo $[\mu \pm 3\sigma]$, significando, obviamente, que a probabilidade média da variável aleatória aumenta à medida que o intervalo alarga. (Ver 9)

Figura 5.3 - Áreas compreendidas pela curva Normal reduzida (%)



b) Distribuição Log - Normal

É uma generalização da distribuição normal, tornada assimétrica por uma mudança da variável, $Y = \ln X$. Se uma variável aleatória Y , tem distribuição normal, então a variável X , diz-se **Log - normal**, e a sua função densidade é,

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \quad x > 0 \quad (5.8)$$

onde, μ_y e σ_y são, respectivamente a média e o desvio padrão da

variável $Y = \ln X$, dados por $\mu_y = \bar{y} = \frac{\sum_{i=1}^n Y_i}{n}$ e

$$\sigma_y = S_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{y})^2}{n-1}}.$$

A distribuição Log - normal ajusta-se bem a variáveis hidrológicas resultantes da multiplicação de muitas variáveis. Isto é, se $X = X_1 X_2 \dots X_n$, então $Y = \ln X = \sum_{i=1}^n \ln X_i = \sum_{i=1}^n Y_i$, que para n grande, tende para a distribuição normal.

c) Distribuição Gama

Diz-se que uma variável aleatória X , tem **distribuição Gama** com parâmetros β e λ , simbolicamente, $G(\beta, \lambda)$, quando a respectiva função densidade de probabilidade é da forma,

$$f(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)} \quad x, \beta, \lambda > 0 \quad (5.9)$$

onde $\Gamma(\beta)$, é a função Gama dada por,

$$\Gamma(\beta) = (\beta - 1)! \quad \beta > 0 \quad (5.10)$$

Os parâmetros β e λ , são determinados, por,

$$\beta = \frac{\bar{x}^2}{S_x^2} = \frac{1}{Cv^2} \quad (5.11)$$

e,

$$\lambda = \frac{\bar{x}}{S_x^2} \quad (5.12)$$

A distribuição Gama é frequentemente aplicada na determinação da probabilidade da precipitação para durações de dias, semanas, meses, e até anos.

d) Distribuição de Pearson tipo III

A distribuição de Pearson tipo III, também chamada distribuição Gama de três parâmetros, introduz um terceiro parâmetro, ε , na distribuição Gama.

Uma variável aleatória X , tem **distribuição Pearson tipo III** com parâmetros β , λ e ε , quando a respectiva função densidade de probabilidade é da forma,

$$f(x) = \frac{\lambda^\beta (x - \varepsilon)^{\beta-1} e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)} \quad x \geq \varepsilon \quad (5.13)$$

Os parâmetros β , λ e ε , podem ser determinados por,

$$\beta = \left(\frac{2}{g} \right)^2, \quad (5.14)$$

$$\lambda = \frac{S_x}{\sqrt{\beta}} \quad (5.15)$$

e

$$\varepsilon = \bar{x} - S_x \sqrt{\beta} \quad (5.16)$$

A distribuição Pearson tipo III é aplicada para descrever a distribuição de probabilidade dos picos de máxima cheia anual.

e) Distribuição Log - Pearson tipo III

Se $Y = \ln X$, segue a distribuição de Pearson tipo III, então diz-se que X tem **distribuição Log - Pearson tipo III**, e tem como função densidade de probabilidade,

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{\Gamma(\beta)} \quad \ln x \geq \varepsilon \quad (5.17)$$

Os parâmetros β , λ e ε , podem ser determinados por,

$$\beta = \left(\frac{2}{g_y} \right)^2, \quad (5.18)$$

$$\lambda = \frac{S_y}{\sqrt{\beta}} \quad (5.19)$$

e

$$\varepsilon = \bar{y} - S_y \sqrt{\beta} \quad (5.20)$$

A distribuição Log - Pearson tipo III é utilizada para descrever a distribuição de frequências dos caudais de cheia.

f) Distribuição Assimptótica de Extremos Tipo I - Gumbel

Também conhecida por distribuição de Gumbel, é bastante aplicada a acontecimentos máximos, por exemplo, a distribuição dos caudais máximos anuais, ou a distribuição das precipitações máximas anuais.

Uma variável aleatória X , **tem distribuição de Gumbel**, com parâmetros α , e u , quando a respectiva função densidade de probabilidade é da forma,

$$f(x) = \frac{1}{\alpha} e^{\left[-\frac{x-u}{\alpha} - e^{\left(-\frac{x-u}{\alpha} \right)} \right]} \quad -\infty < x < +\infty \quad (5.21)$$

e a função distribuição é da forma,

$$F(x) = e^{-e^{\left(-\frac{x-u}{\alpha} \right)}} \quad \alpha > 0 \quad (5.22)$$

Os parâmetros, α e u , podem ser determinados por,

$$\alpha = \frac{\sqrt{6}S_x}{\pi} \quad (5.23)$$

e

$$u = \bar{x} - 0,5772\alpha \quad (5.24)$$

Utilizando a variável reduzida, $y = \frac{x-u}{\alpha}$, vem para a função de distribuição,

$$F(x) = e^{-e^{-y}} \quad (5.25)$$

No Quadro 5.1 apresentam-se, o resumo das distribuições teóricas contínuas de probabilidade mais utilizadas em Hidrologia.

Quadro 5.1 - Distribuições teóricas contínuas de probabilidade utilizadas em Hidrologia.

Distribuição	F. densidade de Probabilidade	Intervalo	Equações dos parâmetros
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$-\infty < x < +\infty$	$\mu = \bar{X}, \sigma = S_x$
Log - Normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}}$ $y = \ln x$	$x > 0$	$\mu_y = \bar{Y}, \sigma_y = S_y$
Gama	$f(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)}$ onde $\Gamma(\beta)$ = função gama	$x > 0$	$\beta = \frac{\bar{X}^2}{S_x^2} = \frac{1}{Cv^2}$ $\lambda = \frac{\bar{X}}{S_x^2}$
Pearson Tipo III	$f(x) = \frac{\lambda^\beta (x - \varepsilon)^{\beta-1} e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)}$	$x \geq \varepsilon$	$\beta = \left(\frac{2}{g}\right)^2, \lambda = \frac{S_x}{\sqrt{\beta}},$ $\varepsilon = \bar{X} - S_x\sqrt{\beta}$
Log - Pearson	$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{\Gamma(\beta)}$ $y = \ln x$	$\ln x \geq \varepsilon$	$\beta = \left(\frac{2}{g_y}\right)^2, \lambda = \frac{S_y}{\sqrt{\beta}},$ $\varepsilon = \bar{Y} - S_y\sqrt{\beta}$
Gumbel	$f(x) = \frac{1}{\alpha} e^{\left[-\frac{x-u}{\alpha} - e^{\left(-\frac{x-u}{\alpha}\right)}\right]}$	$-\infty < x < +\infty$	$\alpha = \frac{\sqrt{6}S_x}{\pi},$ $u = \bar{X} - 0,5772\alpha$

5.3 Exercícios de aplicação

a) Distribuição Normal

Admitindo que a precipitação anual em determinado local, é uma variável aleatória X , com distribuição normal e com parâmetros $\mu = 570mm$ e $\sigma = 120mm$, $N(570;120)$, determinar a **probabilidade** de um valor de precipitação $x \leq 600mm$.

Transformando a variável X com $N(570;120)$ na variável reduzida Z com $N(0,1)$ vem,

$$z = \frac{600 - 570}{120} = 0,25$$

Para obter o correspondente valor de $F(z)$, pode-se recorrer à tabela A9.2:

- Pela tabela, para $z = 0,25$ vem $F(z) = 0,5987$

Isto é probabilidade de a variável X , assumir um valor $x \leq 600\text{mm}$ é de 59,87%.

b) Distribuição Log - Normal

Considerando que o caudal anual de determinado curso de água, é uma variável aleatória X , com distribuição Log - Normal, com $\mu_y = 5,06146$ e $\sigma_y = 0,58906$, determinar a **probabilidade** de se verificar um valor de caudal inferior a $x \leq 150\text{m}^3\text{s}^{-1}$.

Fazendo uma mudança na variável, tal que $Y = \ln X$, vem, $y = \ln x = \ln 150 = 5,01064$, donde a variável reduzida Z é,

$$z = \frac{5,01064 - 5,06146}{0,58906} = -0,09$$

Para obter o correspondente valor de $F(z)$, utiliza-se a tabela A9.2:

- Para $Z = -0,09$ vem, $F(-z) = 1 - F(z) = 1 - 0,5359 = 0,4641$

Que significa que a probabilidade de se verificar um valor de caudal inferior a $x \leq 150\text{m}^3\text{s}^{-1}$ é de 46,41%.

c) Distribuição de Gumbel

Os caudais máximos instantâneos anuais num determinado curso de água seguem a distribuição de Gumbel, com média, $\bar{x} = 226,5m^3s^{-1}$ e desvio padrão, $S = 141,8m^3s^{-1}$. Determine a probabilidade de ocorrer um valor de caudal $x \leq 300m^3s^{-1}$

Os parâmetros, α e u , podem ser determinados pelas equações (5.23) e (5.24),

$$\alpha = \frac{\sqrt{6}S_x}{\pi} = \frac{\sqrt{6} \times 141,8}{\pi} = 110,6$$

$$u = \bar{x} - 0,5772\alpha = 226,5 - 0,5772 \times 110,6 = 162,7$$

$$\text{Utilizando a variável reduzida } y = \frac{x - u}{\alpha} = \frac{300 - 162,7}{110,6} = 1,24141, \text{ a}$$

probabilidade pretendida, pode ser determinada por aplicação da equação (5.25),

$$F(x) = e^{-e^{-y}} = e^{-e^{-1,24141}} = 0,749 = 74,9\%.$$

6. As Distribuição Teóricas e as Variáveis Hidrológicas

Quando se afirma que as variáveis hidrológicas podem ser representadas por algum tipo conhecido de distribuição, não quer dizer que elas sigam perfeitamente essas distribuições teóricas. Obviamente que, quando se trata de variáveis reais, existem **limitações**, que tornam o ajuste perfeito impossível. Por exemplo, como já referido, a precipitação anual é uma variável que segue a distribuição normal. No entanto, a variável aleatória normal, pode assumir qualquer valor no intervalo $(-\infty, +\infty)$, enquanto que a precipitação apenas pode assumir valores positivos ou nulos. Além disso, como se viu, a distribuição normal é uma distribuição simétrica, enquanto que a distribuição de precipitação anual tende a ser assimétrica positiva.

Assim, quando se dispõe de uma amostra de valores de uma determinada variável hidrológica, o objectivo é determinar qual a **distribuição teórica que "melhor" se ajusta à distribuição empírica**. Depois de ajustar um conjunto de variáveis hidrológicas a uma qualquer

distribuição teórica conhecida, grande parte da informação probabilística da amostra pode ser resumida por essa distribuição teórica e pelos respectivos parâmetros.

O ajustamento de um modelo de distribuição à distribuição empírica de variáveis hidrológicas, é habitualmente analisado com base em **testes de hipóteses** estatísticos.

6.1 Testes de Hipóteses

Os **testes de hipóteses** constituem uma área de extrema importância na Estatística Aplicada à Hidrologia. Quando se pretende saber se uma determinada variável aleatória segue uma qualquer distribuição teórica, utiliza-se um teste de hipóteses.

O estabelecimento de um teste de hipóteses consta das seguintes etapas:

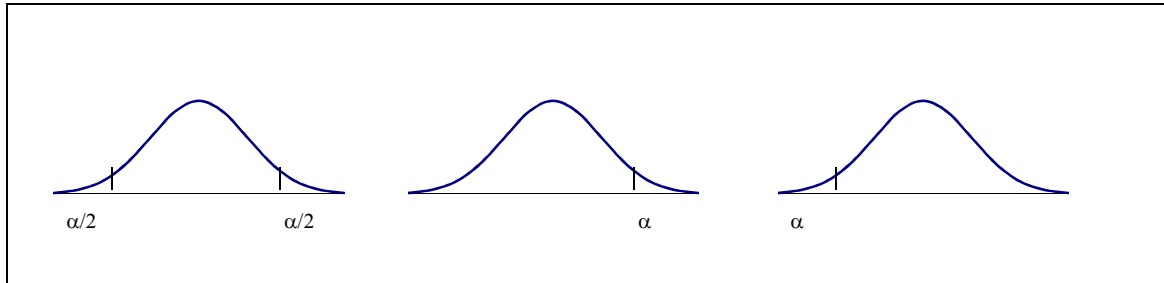
- 1º - Formulação da hipótese a ser testada, H_0 - Hipótese nula
- 2º - Formulação da hipótese alternativa, H_1
- 3º - Selecção da estatística amostral a ser utilizada
- 4º - Estabelecimento da regra de decisão, em função de uma constante c .
- 5º - Selecção do nível de significância, α
- 6º - Utilização da estatística amostral para determinar o valor da constante c , de modo a que, quando H_0 for verdadeira, haja uma probabilidade α de se rejeitar esta hipótese.
- 7º - Rejeição ou aceitação da hipótese H_0 , se a estatística amostral observada cair, respectivamente, na região de rejeição (crítica), ou na região de aceitação.

Ao tomar uma destas duas decisões, pode-se cometer dois tipos de erros: erro de primeira espécie - erro que se comete quando se rejeita H_0 , sendo ela verdadeira e erro de segunda espécie - erro que se comete quando se aceita H_0 , sendo ela falsa.

A probabilidade α de se cometer um erro de primeira espécie, chama-se **nível de significância do teste**. A probabilidade β de se cometer um erro de segunda espécie, chama-se **nível de confiança do teste**. A **região crítica** (Fig. 6.1) do teste é o conjunto dos valores

de uma estatística que determinam a rejeição de H_0 , de acordo com uma regra pré - estabelecida.

Figura 6.1. Diferentes tipos de regiões críticas



De um modo geral, α e β variam em sentido contrário. O que se costuma fazer é fixar α num nível conveniente (5%, 1%, etc.) e procurar, dentro de todas as regiões de nível α , aquela que minimiza β , isto é, aquela que maximiza $1 - \beta$, chamada **potência do teste**. Um melhor teste de nível α é aquele a que corresponde uma maior potência.

a) Teste do Qui-Quadrado

O teste do Qui-Quadrado, χ^2 , é um teste de adequação do ajustamento, onde se pretende determinar se uma dada distribuição teórica é razoável face aos dados disponíveis. Assim, as hipóteses a testar são,

H_0 : A função de distribuição é $F(x)$

H_1 : A função de distribuição não é $F(x)$

O teste do Qui-Quadrado, faz uma comparação entre o número real de observações e o número esperado de observações que caem nas respectivas classes, através do cálculo da estatística,

$$\chi_c^2 = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j} \quad (6.1)$$

que assintoticamente tem distribuição de Qui-Quadrado com $\nu = m - p - 1$ graus de liberdade, sendo m o número de classes, p o número de parâmetros a estimar a partir da amostra, O_j o número de observações na classe j , e E_j o número de observações que seriam de esperar, na classe j , através da distribuição teórica.

A decomposição da amostra em classes, deve ser tal que o efectivo teórico por classe não seja inferior a 5, ou pode ser utilizada a equação (A2.2) de STURGES.

As classes devem ser escolhidas de forma a que cada intervalo de classe corresponda uma probabilidade igual, (classes equiprováveis), donde $E_j = \frac{n}{m}$.

A hipótese H_0 é rejeitada se χ^2_c for maior que $\chi^2_{1-\alpha;v}$ tabelado, para um determinado nível de significância α e v graus de liberdade. (Tabela A9.3).

a) Teste de Kolmogorov - Smirnov

Uma alternativa ao teste do χ^2 , é o teste de Kolmogorov - Smirnov. É um teste, onde não se estimam parâmetros para a distribuição empírica. Para a realização deste teste, deve considerar-se,

1º $F(x)$ a função teórica da distribuição acumulada admitida como hipótese nula, H_0 ;

2º $F_0(x)$ a função de distribuição acumulada para os dados amostrais $\left(\frac{i}{n}\right)$;

3º $D = \max|F(x) - F_0(x)|$, a estatística utilizada;

4º Se, para um determinado nível de significância α , o valor D for maior ou igual ao valor D tabelado (Tabela 9.4), a hipótese H_0 é rejeitada.

7 Análise Frequencial em Hidrologia

Nos sistemas hidrológicos existem muitas vezes eventos extremos, tais como secas ou cheias. O valor de um acontecimento extremo é inversamente proporcional à sua frequência de ocorrência, isto é, um acontecimento extremo ocorre com menos frequência do que um evento moderado. O objectivo da **análise frequencial em hidrologia** é relacionar a magnitude dos valores extremos com a sua frequência de ocorrência, através da utilização de distribuições de probabilidade. Os resultados desta análise podem ser usados em vários problemas de engenharia, tais como, dimensionamento de barragens, pontes, estruturas de controlo de cheias, etc.

Para efectuar a análise frequencial pode-se recorrer ao **posicionamento gráfico dos dados** na forma de uma distribuição cumulativa de probabilidade ou utilizar técnicas analíticas baseadas em **factores de frequência**.

Em qualquer dos casos torna-se necessário introduzir a noção de período de retorno.

7.1 Período de Retorno e Risco Hidrológico

Período de retorno, T , de uma variável X , define-se como o número de anos que deve, em média, decorrer para que o valor dessa variável ocorra ou seja superado.

Designando a probabilidade de a variável aleatória X assumir um valor inferior ou igual a x , por **probabilidade de não excedência, $F(x)$** com, $F(x) = P(X \leq x)$ - Equação (A.1) e designando a probabilidade de a variável aleatória X assumir um valor superior a x , por **probabilidade de excedência $G(x)$** , com $G(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x)$ - Equação (4.2), pode exprimir-se o período de retorno por,

$$T = \frac{1}{G(x)} = \frac{1}{1 - F(x)} \quad (7.1)$$

Risco hidrológico, R , é função do período de retorno e representa a probabilidade de um valor x da variável aleatória X ser excedido em pelo menos uma vez em n anos sucessivos. Exprime-se por,

$$R = \frac{n}{T} = 1 - \left(1 - \frac{1}{T}\right)^n = 1 - (1 - G(x))^n \quad (7.2)$$

7.2 Análise Frequencial por Posicionamento gráfico

Considere-se uma amostra $x_1, x_2, x_3, \dots, x_n$, atribuindo a estes dados amostrais, uma probabilidade empírica $F(x) = P(X \leq x)$ ou $G(x) = P(X > x)$, é possível marcar estes pares de valores $[x, F(x)]$ ou $[x, G(x)]$ em gráficos de modo a poder-se efectuar a análise frequencial. **Este método utiliza-se também para avaliar o ajustamento de uma qualquer distribuição teórica de probabilidade à distribuição empírica dos dados amostrais.**

A função de distribuição de uma determinada distribuição teórica pode ser representada graficamente num **papel de probabilidade** adequado a essa distribuição. Em tal papel, as ordenadas representam os valores da variável X e as abcissas representam a probabilidade de não excedência $F(x) = P(X \leq x)$, a probabilidade de excedência $G(x) = P(X > x)$, o período de retorno T , ou a variável reduzida y . As escalas das ordenadas e das abcissas são feitas de tal modo, que a função de distribuição teórica aparece representada por uma recta. Sendo assim, se os dados amostrais, afectados da respectiva probabilidade empírica, se ajustam à recta da distribuição teórica, então pode-se afirmar que a distribuição empírica segue a distribuição teórica considerada. Nesse caso, é possível efectuar extrapolações para valores extremos.

Suponha-se que se dispõe de todas as observações de uma variável aleatória. Se as n observações (x) forem classificadas por ordem crescente, a **probabilidade empírica** de X tomar valores inferiores ou iguais a um determinado x_i será:

$$F(x) = (P(X \leq x_i)) = \frac{i}{n} \quad (7.3)$$

onde i é o n.º de ordem do valor na amostra. Se as n observações (x) forem classificadas por ordem decrescente, a probabilidade empírica de X tomar valores iguais ou superiores a um determinado x_i será:

$$G(x) = (P(X > x_i)) = \frac{i}{n} \quad (7.4)$$

Neste caso, o menor valor da população teria uma probabilidade igual a zero e o maior valor uma probabilidade igual a um. No entanto, a afectação de probabilidade a uma amostra é mais delicada, pois não há a certeza de que ela contenha o menor e o maior valor da população desconhecida. Das várias fórmulas existentes para afectar cada valor da amostra de uma probabilidade empírica, utilizar-se-á a de WEIBULL, por ser a mais generalizada,

$$F(x) = (P(X \leq x_i)) = \frac{i}{n+1} \quad (7.5)$$

para os n dados classificados por **ordem crescente** e

$$G(x) = (P(X > x_i)) = \frac{i}{n+1} \quad (7.6)$$

para os n dados classificados por **ordem decrescente**.

7.3 Análise Frequencial por Factores de Frequência

A análise frequencial pode ser feita recorrendo a técnicas analíticas baseadas em **factores de frequência**.

CHOW et al (1988) propõe a seguinte fórmula geral para a análise hidrológica de frequências,

$$x_T = \bar{x} + K_T S \quad (7.7)$$

onde, x_T , é o valor do acontecimento associado a determinado período de retorno, K_T , é o **factor de frequência** que é função do período de retorno, T , e do tipo de distribuição de probabilidade a ser utilizada na análise.

Se a variável em análise é $y = \ln x$, o mesmo método pode ser utilizado, aplicado aos logaritmos dos dados,

$$y_T = \bar{y} + K_T S_y \quad (7.8)$$

O factor de frequência proposto por VEN TE CHOW é aplicável a muitas distribuições de probabilidade utilizadas na análise hidrológica de frequências. Para uma determinada distribuição teórica, é possível determinar uma relação, $K - T$, entre o factor de frequência e o correspondente período de retorno, relação esta que pode ser expressa por tabelas ou em termos matemáticos.

Para determinar o valor de x_T (Equação 7.7), é então necessário calcular os parâmetros estatísticos para a distribuição proposta e determinar para um dado período de retorno, o factor de frequência.

Seguidamente descreve-se a relação teórica $K - T$, para várias distribuições de probabilidade.

a) Distribuição Normal

O factor de frequência pode ser expresso por,

$$K_T = \frac{x_T - \mu}{\sigma} = z \quad (7.9)$$

que é a mesma expressão da variável normal reduzida Z , definida na equação (5.5), e que se encontra tabelada (Tabela A9.2).

b) Distribuição Log - Normal

Para a distribuição Log - Normal O factor de frequência pode ser expresso por,

$$K_T = \frac{Y_T - \mu_y}{\sigma_y} \quad (7.10)$$

onde $y = \ln x$. Este factor de frequência aplica-se à equação (7.8)

c) Distribuição Pearson Tipo III

O factor de frequência para esta distribuição, é expresso por meio de uma tabela em função do coeficiente de assimetria, g , e do período de retorno, T . (Tabela 9.5)

d) Distribuição Log - Pearson Tipo III

Igual ao caso anterior, no entanto, o factor de frequência obtido pela tabela deverá ser aplicado à equação (7.8).

e) Distribuição de Gumbel (Extremos tipo I)

Para esta distribuição, o factor de frequência é determinado por,

$$K_T = -\frac{\sqrt{6}}{\pi} \left\{ 0,5772 + \ln \left[\ln \left(\frac{T}{T-1} \right) \right] \right\} \quad (7.11)$$

Para expressar T , em termos de K_T , utiliza-se a seguinte equação,

$$T = -\frac{1}{1 - e^{-e^{-\left[0,5772 + \frac{\pi K_T}{\sqrt{6}} \right]}}} \quad (7.12)$$

8 Exercícios de Aplicação**8.1 Ajustamento de uma distribuição empírica à distribuição Normal**

Verificar o ajustamento das precipitações anuais ocorridas na estação meteorológica de Castro D'Aire (Quadro 2.1) à distribuição Normal.

Esta verificação pode ser feita de duas maneiras: por **posicionamento gráfico dos dados** ou através de um **teste de adequação do ajustamento**.

a) Posicionamento gráfico (Ver ponto 7.2)

A função de distribuição da distribuição Normal pode ser representada graficamente num **papel de probabilidade Normal**. Em tal papel, as ordenadas representam os valores da variável X e as abcissas representam a **probabilidade** $F(x) = P(X \leq x)$ ou $G(x) = P(X > x)$. As escalas das ordenadas e das abcissas são feitas de tal modo, que a função de distribuição teórica aparece representada por uma recta. Assim, num papel de probabilidade normal, **qualquer distribuição normal** terá como gráfico uma linha recta, correspondendo a média dessa distribuição ao ponto 50% e um desvio padrão para cada lado da média, aos pontos 15.87% e 84.13%, respectivamente (ver Figura 5.2 e 5.3).

Neste caso² a recta da distribuição normal teórica desenha-se no papel normal unindo os três pares de pontos,

$$(\bar{x} - S; 15,87\%) = (1193,1; 15,87\%)$$

$$(\bar{x}; 50\%) = (1672,5; 50\%)$$

$$(\bar{x} + S;) = (2151,9; 84,13\%)$$

Esta recta corresponde à distribuição Normal teórica, se os valores da amostra, afectados da respectiva probabilidade empírica, ajustarem à recta, então pode-se afirmar que a série de precipitações anuais segue a distribuição Normal.

Para atribuir uma probabilidade empírica aos valores da amostra, utiliza-se a expressão (7.5), que dá a probabilidade de não excedência, $F(x)$, para os n valores da amostra, ordenados de forma crescente. Quadro 8.1.

A recta teórica de probabilidade Normal e os valores da distribuição empírica da precipitação anual estão representados na

² A média e o desvio padrão foram determinados no Exemplo 3.1 e 3.2 e são, respectivamente $\bar{x} = 1672,5mm$ e $S = 479,4mm$

Figura 8.1, onde se pode verificar o ajustamento à recta, donde se pode afirmar que **a série de precipitações em estudo tem distribuição Normal.**

Quadro 8.1 Probabilidade de não excedência, $F(x)$, para os valores de precipitação anual em Castro Daire.

Ano	Prec. (x_i)	Prec.ordenada (x_i)	i	F(x)
1916/17	2118,2	870,9	1	1,3
1917/18	1001,2	903,5	2	2,5
1918/19	2093,2	912,8	3	3,8
1919/20	1556,4	915,8	4	5,0
1920/21	1290,6	1001,2	5	6,3
1921/22	1785,4	1039,2	6	7,5
1922/23	1830,2	1055,4	7	8,8
1923/24	2150,2	1076,1	8	10,0
1924/25	1749,6	1127,1	9	11,3
1925/26	2221,6	1144,5	10	12,5
1926/27	2024,1	1180,0	11	13,8
1927/28	1923,7	1201,0	12	15,0
1928/29	1127,1	1239,6	13	16,3
1929/30	2630,9	1247,2	14	17,5
1930/31	1481,2	1254,0	15	18,8
1931/32	1461,0	1275,7	16	20,0
1932/33	1334,4	1290,6	17	21,3
1933/34	1301,2	1298,7	18	22,5
1934/35	1581,0	1300,3	19	23,8
1935/36	3249,6	1301,2	20	25,0
1936/37	2069,0	1334,4	21	26,3
1937/38	1254,0	1344,7	22	27,5
1938/39	1974,0	1392,9	23	28,8
1939/40	2059,6	1411,7	24	30,0
1940/41	2569,6	1422,9	25	31,3
1941/42	1520,6	1426,8	26	32,5
1942/43	1664,2	1432,0	27	33,8
1943/44	1344,7	1442,0	28	35,0
1944/45	915,8	1451,9	29	36,3
1945/46	1763,0	1461,0	30	37,5
1946/47	2079,3	1478,1	31	38,8
1947/48	1411,7	1481,2	32	40,0
1948/49	912,8	1496,4	33	41,3
1949/50	1201,0	1504,2	34	42,5
1950/51	1903,9	1520,6	35	43,8
1951/52	1625,0	1556,4	36	45,0
1952/53	1076,1	1567,9	37	46,3
1953/54	1275,7	1578,2	38	47,5
1954/55	1699,5	1581,0	39	48,8
1955/56	2150,9	1585,4	40	50,0
1956/57	1039,2	1588,2	41	51,3

Quadro 8.1 (cont.) Probabilidade de não excedência, $F(x)$, para os valores de precipitação anual em Castro Daire.

Ano	Prec. (x_i)	Prec.ordenada (x_i)	i	F(x)
1957/58	1588,2	1595,9	42	52,5
1958/59	1746,1	1603,3	43	53,8
1959/60	2563,6	1625,0	44	55,0

1960/61	1987,4	1664,2	45	56,3
1961/62	1585,4	1689,7	46	57,5
1962/63	1832,1	1699,5	47	58,8
1963/64	2201,2	1746,1	48	60,0
1964/65	1180,0	1749,6	49	61,3
1965/66	2806,9	1763,0	50	62,5
1966/67	1595,9	1785,4	51	63,8
1967/68	1422,9	1826,1	52	65,0
1968/69	2280,0	1830,2	53	66,3
1969/70	1496,4	1832,1	54	67,5
1970/71	1567,9	1903,9	55	68,8
1971/72	1300,3	1923,7	56	70,0
1972/73	1478,1	1930,1	57	71,3
1973/74	1689,7	1974,0	58	72,5
1974/75	1239,6	1987,4	59	73,8
1975/76	903,5	2000,2	60	75,0
1976/77	2314,0	2024,1	61	76,3
1977/78	2124,2	2059,6	62	77,5
1978/79	2599,1	2069,0	63	78,8
1979/80	1451,9	2079,3	64	80,0
1980/81	1144,5	2093,2	65	81,3
1981/82	1504,2	2118,2	66	82,5
1982/83	1426,8	2124,2	67	83,8
1983/84	1603,3	2150,2	68	85,0
1984/85	2000,2	2150,9	69	86,3
1985/86	1578,2	2201,2	70	87,5
1986/87	1392,9	2221,6	71	88,8
1987/88	1930,1	2280,0	72	90,0
1988/89	870,9	2314,0	73	91,3
1989/90	1432,0	2563,6	74	92,5
1990/91	1442,0	2569,6	75	93,8
1991/92	1055,4	2599,1	76	95,0
1992/93	1247,2	2630,9	77	96,3
1993/94	1826,1	2806,9	78	97,5
1994/95	1298,7	3249,6	79	98,8

b) Teste do Qui-Quadrado, χ^2 (Ver ponto 6.1)

Para melhor ajuizar da qualidade do ajustamento da distribuição normal à distribuição empírica de precipitações anuais, utiliza-se o teste de hipótese do χ^2 .

As hipóteses a testar são,

Figura 8.1. Distribuição das Precipitações anuais

H_0 : A função de distribuição é normal

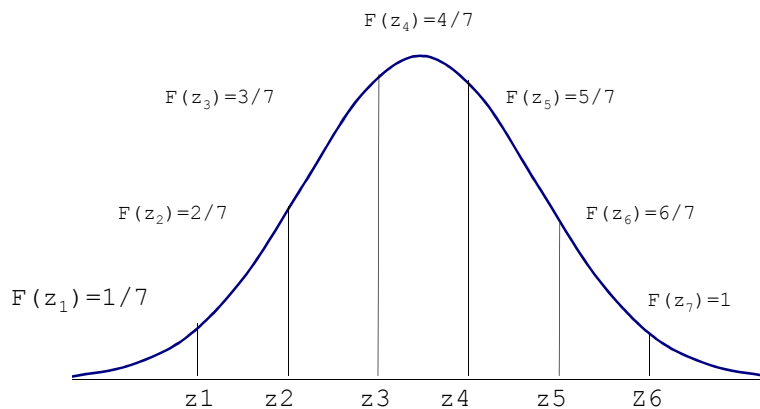
H_1 : A função de distribuição não é normal

O número de classes, m , para esta amostra é 7 (determinado no Exemplo 2.1)

Uma vez que é necessário trabalhar com as tabelas para a distribuição Normal, utilizar-se-á a variável reduzida z .

Como as classes devem ser equiprováveis vem para a probabilidade de cada classe,

$$F(z) = \frac{1}{7} = 0,1428$$



Os z_i serão calculados, a partir dos valores $F(z_i)$ conhecidos, por consulta da tabela 9.2. A partir de z_i determina-se facilmente os intervalos das classes, x_i , sabendo que $\bar{X} = 1672,5mm$ e $S = 479,4mm$. Como se mostra no Quadro 8.2.

Quadro 8.2 Cálculo dos intervalos e limites das classes

z_i	$F(z_i)$	z_i	$x_i = z_i \times S + \bar{X}$
z_1	$1/7 = 0,1429$	$-1,0674$	$1160,8$
z_2	$2/7 = 0,2857$	$-0,5659$	$1401,2$
z_3	$3/7 = 0,4286$	$-0,18$	$1586,2$
z_4	$4/7 = 0,5714$	$0,18$	$1758,8$
z_5	$5/7 = 0,7143$	$0,5659$	$1943,8$
z_6	$6/7 = 0,8571$	$1,0674$	$2184,2$

Com estes elementos pode-se construir o Quadro 8.3 e calcular o

$$\chi_c^2 = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j} \quad (\text{Equação 6.1}).$$

Quadro 8.3 Teste do Qui-quadrado

Lim. de Classe	Nº de elementos esperados em cada classe (E_j)	Nº de elementos observados em cada classe (O_j)	$\frac{(O_j - E_j)^2}{E_j}$
$\leq 1160,8$	11,2857	10	0,1465
1160,8 - 1401,2	11,2857	13	0,2604
1401,2 - 1586,2	11,2857	17	2,8933
1586,2 - 1758,8	11,2857	9	0,4629
1758,8 - 1943,8	11,2857	8	0,9566
1943,8 - 2184,2	11,2857	12	0,0452
$> 2184,2$	11,2857	10	0,1465
TOTAL	79	79	4,9114

Da tabela 9.3 vem, para $\alpha = 0,05$ e $\nu = m - p - 1 = 7 - 2 - 1 = 4$ graus de liberdade,

$$\chi_{0,95;4}^2 = 9,49$$

Como $\chi_c^2 = 4,9114 < \chi_{0,95;4}^2 = 9,49$ pode-se dizer que a **hipótese de normalidade não é rejeitada**, o que vem confirmar a análise gráfica feita na alínea anterior.

8.2 Ajustamento de uma distribuição empírica à distribuição de Gumbel

Verificar o ajustamento das precipitações diárias máximas anuais ocorridas na estação meteorológica de Castro D'Aire (Quadro 8.4) à distribuição de Gumbel.

A verificação do ajustamento irá ser realizada de duas formas: por **posicionamento gráfico dos dados** ou através de um **teste de adequação do ajustamento**.

a) Posicionamento gráfico (Ver ponto 7.2)

Tal como a distribuição Normal, também a distribuição de Gumbel pode ser representada por uma recta quando desenhada no **papel de Gumbel**. Neste papel, as ordenadas representam os valores da variável X

e as abcissas representam a **probabilidade** $F(x) = P(X \leq x)$ e a **variável reduzida** $y = \frac{x - u}{\alpha}$. Para traçado da recta basta unir, por exemplo, três pontos (x_i, y_i) escolhidos, com $x_i = u + y_i \times \alpha$. Para tal é necessário determinar os parâmetros α e u , que como já se viu são determinados em função da média e do desvio padrão da amostra (Equações A5.23 e A5.24).

A média e o desvio padrão das precipitações diárias máximas anuais em Castro D'Aire, são respectivamente $\bar{x} = 89,6\text{mm}$ e $S = 24,9\text{mm}$, donde os parâmetros são, $\alpha = \frac{\sqrt{6} \times 24,9}{\pi} = 19,4$ e $u = 89,6 - 0,5772 \times 19,4 = 78,4$.

Para traçado da recta teórica, basta atribuir valores a y , obter os correspondentes valores de x e marcar estes pares de valores no papel de Gumbel. Por exemplo,

y_i	$x_i = u + y_i \times \alpha$
-1	59,0
0	78,4
1	97,8

Com os pares $(59,0;-1)$, $(78,4;0)$ e $(97,8;1)$ desenha-se a recta da Figura 8.2.

Para atribuir uma probabilidade empírica aos valores da amostra, utiliza-se a expressão (7.5), que dá a probabilidade de não excedência, $F(x)$, para os n valores da amostra, ordenados de forma crescente. Estes valores, apresentados no Quadro 8.4, foram marcados no papel de Gumbel (Fig. 8.2), onde se pode verificar o ajustamento à recta teórica, donde se pode afirmar que **a série de precipitações em estudo segue a distribuição de Gumbel**.

Quadro 8.4 Precipitação diária máxima anual (mm) em Castro D'Aire e Probabilidade de não excedência, $F(x)$.

Ano	Prec. (x_i)	Prec.ordenada (x_i)	i	F(x)
1916/17	199,4	49,6	1	1,3
1917/18	49,6	52,6	2	2,5
1918/19	120,4	53,3	3	3,8
1919/20	105,0	53,4	4	5,0
1920/21	73,6	54,3	5	6,3
1921/22	72,4	59,1	6	7,5

Quadro 8.4 (Cont.) Precipitação diária máxima anual (mm) em Castro D'Aire e Probabilidade de não excedência, $F(x)$

Ano	Prec. (x_i)	Prec.ordenada (x_i)	i	F(x)
1922/23	99,6	62,3	7	8,8
1923/24	79,8	64,6	8	10,0
1924/25	98,6	65,6	9	11,3
1925/26	102,0	66,5	10	12,5
1926/27	82,0	67,3	11	13,8
1927/28	99,3	67,4	12	15,0
1928/29	52,6	69,7	13	16,3
1929/30	101,2	71,0	14	17,5
1930/31	98,2	71,2	15	18,8
1931/32	77,8	72,2	16	20,0
1932/33	53,4	72,4	17	21,3
1933/34	65,6	72,4	18	22,5
1934/35	100,8	72,6	19	23,8
1935/36	105,8	73,6	20	25,0
1936/37	101,6	73,6	21	26,3
1937/38	82,6	73,8	22	27,5
1938/39	161,4	74,5	23	28,8
1939/40	72,4	74,9	24	30,0
1940/41	130,8	75,1	25	31,3
1941/42	84,9	75,2	26	32,5
1942/43	111,2	75,8	27	33,8
1943/44	124,0	77,4	28	35,0
1944/45	83,4	77,8	29	36,3
1945/46	73,6	78,4	30	37,5
1946/47	78,4	78,8	31	38,8
1947/48	99,6	79,1	32	40,0
1948/49	64,6	79,8	33	41,3
1949/50	80,4	80,4	34	42,5
1950/51	78,8	82,0	35	43,8
1951/52	99,0	82,6	36	45,0
1952/53	90,6	83,4	37	46,3
1953/54	93,0	83,6	38	47,5
1954/55	71,0	83,7	39	48,8
1955/56	121,2	84,4	40	50,0
1956/57	54,3	84,9	41	51,3
1957/58	104,4	84,9	42	52,5
1958/59	88,6	86,6	43	53,8
1959/60	84,9	88,5	44	55,0
1960/61	86,6	88,6	45	56,3
1961/62	59,1	90,2	46	57,5
1962/63	72,6	90,6	47	58,8
1963/64	118,8	91,1	48	60,0
1964/65	90,2	92,5	49	61,3
1965/66	111,0	93,0	50	62,5
1966/67	140,6	93,1	51	63,8
1967/68	83,7	96,8	52	65,0
1968/69	67,4	98,2	53	66,3
1969/70	84,4	98,6	54	67,5
1970/71	66,5	99,0	55	68,8
1971/72	69,7	99,3	56	70,0
1972/73	96,8	99,6	57	71,3

Quadro 8.4 (Cont.) Precipitação diária máxima anual (mm) em Castro D'Aire e Probabilidade de não excedência, $F(x)$

Ano	Prec. (x_i)	Prec.ordenada (x_i)	i	F(x)
1973/74	74,9	99,6	58	72,5
1974/75	91,1	100,8	59	73,8
1975/76	73,8	101,2	60	75,0
1976/77	83,6	101,6	61	76,3
1977/78	125,5	102,0	62	77,5
1978/79	119,3	104,4	63	78,8
1979/80	75,8	105,0	64	80,0
1980/81	79,1	105,8	65	81,3
1981/82	92,5	111,0	66	82,5
1982/83	93,1	111,2	67	83,8
1983/84	77,4	113,4	68	85,0
1984/85	88,5	117,5	69	86,3
1985/86	74,5	118,8	70	87,5
1986/87	72,2	119,3	71	88,8
1987/88	117,5	120,4	72	90,0
1988/89	75,1	121,2	73	91,3
1989/90	71,2	124,0	74	92,5
1990/91	62,3	125,5	75	93,8
1991/92	67,3	130,8	76	95,0
1992/93	75,2	140,6	77	96,3
1993/94	113,4	161,4	78	97,5
1994/95	53,3	199,4	79	98,8
$\bar{x} =$	89,6	89,6		
$S =$	24,9	24,9		
$\alpha =$	19,4	19,4		
$u =$	78,4	78,4		

b) Teste de Kolmogorov - Smirnov (Ver ponto 6.1)

Para melhor ajuizar da qualidade do ajustamento da distribuição de Gumbel à distribuição empírica de precipitações máximas anuais, utilizar-se-á o teste de Kolmogorov - Smirnov, seguindo os passos descritos na alínea b) do ponto 6.1:

1º Admite-se que a Função de distribuição de Gumbel,

$$F(x) = e^{-e^{\left(-\frac{x-u}{\alpha}\right)}} = e^{-e^{\left(-\frac{x-78,4}{19,4}\right)}} \text{ é a hipótese nula, } H_0;$$

2º Considera-se que $F_0(x) = \frac{i}{n} = \frac{i}{79}$ é a função de distribuição para os valores da amostra;

3º Calcula-se a estatística $D = \max|F(x) - F_0(x)|$;

4º Rejeita-se H_0 , se para um nível de significância $\alpha = 0,05$, o valor de D for maior ou igual ao valor D tabelado (Tabela 9.4).

Figura 8.2. Distribuição das Precipitações diárias máximas anuais

No Quadro 8.5 mostram-se os passos necessários para efectuar este teste, já explicado no ponto 6.1.

Pela análise do Quadro 8.5, pode-se concluir que a **hipótese nula não é rejeitada**, uma vez que o $\max|F(x) - F_0(x)| = 0,0545$ é inferior ao indicado na tabela 9.4 - $\left(\frac{1,36}{\sqrt{79}} = 0.1125\right)$, o que vem confirmar a análise gráfica feita na alínea anterior.

Quadro 8.5 Teste de Kolmogorov - Smirnov para ajustamento das precipitações diárias máximas anuais (mm) à distribuição de Gumbel.

i	x_i ordenada	$y_i = \frac{x_i - 78,4}{19,4}$	$F_0(x) = \frac{i}{79}$	$F(x) = e^{-e^{(-y_i)}}$	$ F(x) - F_0(x) $
1	49,6	-1,48	0,013	0,012	0,0005
2	52,6	-1,33	0,025	0,023	0,0024
3	53,3	-1,29	0,038	0,026	0,0118
4	53,4	-1,29	0,051	0,027	0,0239
5	54,3	-1,24	0,063	0,031	0,0318
6	59,1	-0,99	0,076	0,067	0,0088
7	62,3	-0,83	0,089	0,101	0,0127
8	64,6	-0,71	0,101	0,131	0,0296
9	65,6	-0,66	0,114	0,145	0,0310
10	66,5	-0,61	0,127	0,158	0,0316
11	67,3	-0,57	0,139	0,170	0,0312
12	67,4	-0,57	0,152	0,172	0,0201
13	69,7	-0,45	0,165	0,209	0,0448
14	71,0	-0,38	0,177	0,232	0,0545
15	71,2	-0,37	0,190	0,235	0,0453
16	72,2	-0,32	0,203	0,253	0,0504
17	72,4	-0,31	0,215	0,257	0,0414
18	72,4	-0,31	0,228	0,257	0,0287
19	72,6	-0,30	0,241	0,260	0,0196
20	73,6	-0,25	0,253	0,278	0,0252
21	73,6	-0,25	0,266	0,278	0,0125
22	73,8	-0,24	0,278	0,282	0,0036
23	74,5	-0,20	0,291	0,295	0,0038
24	74,9	-0,18	0,304	0,302	0,0014
25	75,1	-0,17	0,316	0,306	0,0103
26	75,2	-0,16	0,329	0,308	0,0211
27	75,8	-0,13	0,342	0,319	0,0225
28	77,4	-0,05	0,354	0,349	0,0050
29	77,8	-0,03	0,367	0,357	0,0100
30	78,4	0,00	0,380	0,368	0,0113
31	78,8	0,02	0,392	0,376	0,0164
32	79,1	0,04	0,405	0,382	0,0234
33	79,8	0,07	0,418	0,395	0,0228

Quadro 8.5 (Cont.) Teste de Kolmogorov - Smirnov para ajustamento das precipitações diárias máximas anuais (mm) à distribuição de Gumbel.

i	x_i ordenada	$y_i = \frac{x_i - 78,4}{19,4}$	$F_0(x) = \frac{i}{79}$	$F(x) = e^{-e^{-y_i}}$	$ F(x) - F_0(x) $
34	80,4	0,10	0,430	0,406	0,0241
35	82,0	0,19	0,443	0,436	0,0067
36	82,6	0,22	0,456	0,447	0,0082
37	83,4	0,26	0,468	0,462	0,0061
38	83,6	0,27	0,481	0,466	0,0151
39	83,7	0,27	0,494	0,468	0,0259
40	84,4	0,31	0,506	0,481	0,0258
41	84,9	0,34	0,519	0,490	0,0294
42	84,9	0,34	0,532	0,490	0,0421
43	86,6	0,42	0,544	0,520	0,0245
44	88,5	0,52	0,557	0,553	0,0044
45	88,6	0,53	0,570	0,554	0,0154
46	90,2	0,61	0,582	0,581	0,0015
47	90,6	0,63	0,595	0,587	0,0077
48	91,1	0,66	0,608	0,595	0,0124
49	92,5	0,73	0,620	0,617	0,0031
50	93,0	0,75	0,633	0,625	0,0082
51	93,1	0,76	0,646	0,626	0,0193
52	96,8	0,95	0,658	0,679	0,0211
53	98,2	1,02	0,671	0,698	0,0269
54	98,6	1,04	0,684	0,703	0,0194
55	99,0	1,06	0,696	0,708	0,0118
56	99,3	1,08	0,709	0,712	0,0029
57	99,6	1,09	0,722	0,716	0,0060
58	99,6	1,09	0,734	0,716	0,0186
59	100,8	1,16	0,747	0,730	0,0168
60	101,2	1,18	0,759	0,735	0,0247
61	101,6	1,20	0,772	0,739	0,0328
62	102,0	1,22	0,785	0,744	0,0409
63	104,4	1,34	0,797	0,770	0,0275
64	105,0	1,37	0,810	0,776	0,0340
65	105,8	1,41	0,823	0,784	0,0386
66	111,0	1,68	0,835	0,830	0,0051
67	111,2	1,69	0,848	0,832	0,0162
68	113,4	1,81	0,861	0,848	0,0123
69	117,5	2,02	0,873	0,875	0,0020
70	118,8	2,08	0,886	0,883	0,0030
71	119,3	2,11	0,899	0,886	0,0129
72	120,4	2,17	0,911	0,892	0,0196
73	121,2	2,21	0,924	0,896	0,0281
74	124,0	2,35	0,937	0,909	0,0275
75	125,5	2,43	0,949	0,916	0,0337
76	130,8	2,70	0,962	0,935	0,0268
77	140,6	3,21	0,975	0,960	0,0143
78	161,4	4,28	0,987	0,986	0,0011
79	199,4	6,24	1,000	0,998	0,0019

8.3 Análise Frequencial

8.3.1 Relativamente às precipitações anuais em Castro D'Aire (Exercício 8.1), determinar: a precipitação associada a um período de retorno de 100 anos; o período de retorno do maior valor de precipitação.

Depois de se ter verificado (Exercício 8.1) que as precipitações anuais em Castro D'Aire seguem a distribuição Normal é possível efectuar a análise frequencial pretendida. Para esta a análise pode-se recorrer ao **posicionamento gráfico dos dados** na forma de uma distribuição cumulativa de probabilidade ou utilizar técnicas analíticas baseadas em **factores de frequência**.

a) Análise Frequencial por Posicionamento gráfico

Utilizando a Equação (7.1) é possível determinar a probabilidade de não excedência correspondente a um **período de retorno igual a 100 anos**,

$$T = \frac{1}{1 - F(x)} \Leftrightarrow F(x) = 1 - \frac{1}{T} = 1 - \frac{1}{100} = 0,99 = 99\%$$

Com este valor é possível tirar da recta teórica Normal, o correspondente valor de X . Para $F(x) = 99\%$ vem que $x \approx 2770\text{mm}$ (Ver Figura A8.1). Isto é a **precipitação associada a um $T = 100$ anos, é 2770 mm**.

O **maior valor de precipitação** anual em Castro D'Aire é $x = 3249,6\text{mm}$ (Ver Quadro A8.1). Com este valor pode-se ler na recta teórica o correspondente valor de $F(x)$. Pela leitura da Figura A8.1. vem, para $x = 3249,6\text{mm}$ um valor de $F(x) \approx 99,95\%$.

Donde, o **período de retorno de um valor de precipitação 3249,6mm** é,

$$T = \frac{1}{1 - F(x)} = \frac{1}{1 - 0,9995} = 2000 \text{ anos.}$$

b) Análise Frequencial por Factores de Frequência

Sabendo que o factor de frequência, K_T , para a distribuição Normal, é igual à variável reduzida z , (Equação 7.9), a Equação (7.7) transforma-se em,

$$x_T = \bar{x} + zS \quad (7.13)$$

que para a distribuição em estudo é,

$$x_T = 1672,5 + z \times 479,4 \quad (7.14)$$

Um **período de retorno igual a 100 anos** corresponde a uma probabilidade de não excedência de 99%. Consultando a tabela 9.2, vem para $F(z) = 0,99$ um valor de $z = 2,33$, donde da equação (7.14) vem que a **precipitação associada a um $T = 100$ anos é,**

$$x_T = 1672,5 + 2,33 \times 479,4 = 2789,5 \text{ mm}$$

O **maior valor de precipitação** anual em Castro D'Aire é $x = 3249,6 \text{ mm}$. Resolvendo a equação (A7.14) em ordem a z , vem,

$$3249,6 = 1672,5 + z \times 479,4 \Leftrightarrow z = \frac{3249,6 - 1672,5}{479,4} = 3,29$$

Pela Tabela A9.2, para $z = 3,29$ vem $F(z) = 0,9995$, e o **período de retorno do valor 3249,6 mm é,**

$$T = \frac{1}{1 - F(x)} = \frac{1}{1 - 0,9995} = 2000 \text{ anos.}$$

8.3.2 Relativamente às precipitações diárias máximas anuais em Castro D'Aire (Exercício 8.2), determinar: a precipitação diárias máxima associada a um período de retorno de 100 anos; o período de retorno do maior valor de precipitação.

Depois de se ter verificado (Exercício 8.2) que as precipitações diárias máximas anuais em Castro D'Aire seguem a distribuição de Gumbel é possível efectuar a análise frequencial pretendida.

a) Análise Frequencial por Posicionamento gráfico

Pela recta teórica da distribuição de Gumbel (Figura 8.2), vem **para $T = 100$ anos, uma precipitação diária máxima anual de aproximadamente 170 mm.**

O maior valor de precipitação diária máxima anual em Castro D'Aire é $x = 199,4\text{mm}$ (Ver Quadro 8.4). Com este valor pode-se ler na recta teórica o correspondente valor de T . Pela leitura da Figura A8.2. vem, para $x = 199,4\text{ mm}$ um valor de $T \approx 450$ anos.

b) Análise Frequencial por Factores de Frequência

Para esta distribuição, a equação (7.7) é,

$$x_T = 89,6 + K_T \times 24,9 \quad (7.15)$$

Sabendo que o factor de frequência, K_T , para a distribuição de Gumbel, dado pela Equação (A7.11), vem para um período de retorno de 100 anos,

$$K_T = -\frac{\sqrt{6}}{\pi} \left\{ 0,5772 + \ln \left[\ln \left(\frac{100}{100-1} \right) \right] \right\} = 3,137$$

Por (7.15) vem uma precipitação diária máxima anual de,

$$x_T = 89,6 + 3,137 \times 24,9 = 167,7\text{mm}$$

O maior valor de precipitação diária máxima anual é, $x = 199,4\text{mm}$, resolvendo a Equação (7.15), em ordem a K_T vem,

$$199,4 = 89,6 + K_T \times 24,9 \Leftrightarrow K_T = \frac{199,4 - 89,6}{24,9} = 4,410$$

Para determinar o período de retorno do, maior valor de precipitação diária máxima anual, basta resolver a Equação (7.12)

$$T = -\frac{1}{1 - e^{-e^{-\left[0,5772 + \frac{\pi K_T}{\sqrt{6}}\right]}}} = -\frac{1}{1 - e^{-e^{-\left[0,5772 + \frac{\pi \times 4,410}{\sqrt{6}}\right]}}} = 510 \text{ anos.}$$